Global Maritime Forum Workshop
sponsored by
The National Maritime
Intelligence-Integration Office (NMIO)
and The NASA Ames Research Center

Turning the Corner in the Maritime Domain

Leveraging Data to Achieve Effective Understanding

REPORT ON THE WORKSHOP

NASA Ames Research Center
Moffett Field, California
2-3 June 2015

Global Maritime Forum Workshop
"Turning the Corner in the Maritime Domain -
Leveraging Data to Achieve
Effective Understanding"

Report on the workshop jointly sponsored by
The National Maritime
Intelligence-Integration Office (NMIO)
and
The NASA Ames Research Center

June 2–3, 2015
Moffett Field, California

Table of Contents

This report summarizes the presentations of the GMF workshop as interpreted by Dr. Paul Shapiro, Chief Science and Technology Advisor, National Maritime Intelligence-Integration Office.

The conference adheres to a variation of the Chatham House Rule. Accordingly, beyond the points expressed in the presentations, no attributions have been included in this conference report.

Executive Summary

Rear Admiral (RADM) Elizabeth Train of the National Maritime Intelligence-Integration Office (NMIO) and Dr. Eugene Tu of the NASA Ames Research Center at NASA Ames, Moffett Field, California hosted the 2015 Global Maritime Forum (GMF) Workshop on 2–3 June. Convened in the heart of Silicon Valley, the workshop fostered participation from the Global Maritime Community of Interest (GMCOI) and experts, from sources as diverse as numerous startups companies, to major internet powerhouses such as Google and Amazon, to representatives from academic institutions such as Stanford University and Cornell University, as well as non-government organizations (NGOs) such as Pew Charitable Trusts and The World Bank, and NASA scientists specialized in the use of high performance computing, i.e., supercomputing. The two-day event highlighted the latest developments in high performance computing on the first day and the latest data analytics and platforms in the maritime domain on the second day.

The Day 1 keynote by Dr. Eng Lim Goh, Chief Technology Officer of the Silicon Graphics International Corporation, set the stage with the practical applications and advances in high performance computing. Dr. Goh introduced two approaches or concepts to computing vast quantities of information that pervaded discussions throughout the workshop: one, starting with a simple question or input and relating all known data to output extensive information (expansive approach), and the other, reducing an immense amount of data to the simplest possible output.

The first-day's sessions and speakers following Dr. Goh dove into the technical details and illustrated these concepts. Discussion on data platforms included Amazon Web Services, the startup Planet OS and their aggregation of tens of thousands of data streams in partnership with National Oceanic and Atmospheric Administration (NOAA), and the newly-acquired Google Skybox providing an end-to-end solution of satellite imagery and a big data analytics computing platform. The specific data analytic techniques highlighted included the entity resolution work of Novetta Solutions and the data mining and machine learning being applied at NASA for satellite imagery classification and safety of flight for the Federal Aviation Administration (FAA). Connecting the platforms and analytics techniques to the operator, the final session of the day focused on data visualization and accessibility and introduced another important theme, "The critical goal of work on big data and machine intelligence should be teaming computers with humans to take advantage of the unique problem solving capabilities of each."

The Day 2 keynote by Mr. Kshemendra Paul, Program Manager for the Information Sharing Environment (PM-ISE), provided a vision of enhancing national security through responsible information sharing, and was followed by sessions highlighting the wide ranging work being done internationally and commercially to support maritime missions of common concern. The first session of the day focused on illegal, unreported, and unregulated (IUU) fishing with presentations from Google highlighting its strong partnerships with the United States and international governments to advance its Google Oceans platform, algorithms being developed to identify fishing behavior, and NOAA's capabilities for deploying unmanned aerial vehicles. The second session on interdiction was chaired by the director of the Global Maritime Operation Threat Response (MOTR) Coordination Center, Mr. Scott Genovese, and provided insights from the NATO Center for Maritime Research and Experimentation (CMRE), the European Commission Joint Research Centre (EU JRC), and the United Nations (UN) Panel of Experts (PoE) monitoring the sanctions against the Democratic People's Republic of Korea. The last session of the workshop chaired by Mr. Michael Rodriguez, Deputy Administrator of the Department of Transportation Maritime Administration (MARAD), highlighted the efforts by firms to analyze the global supply chain and support efficient and effective commercial operations in the maritime domain.

While the GMF workshop had particularly strong presenters and relevant content, the event also took advantage of the equally strong participants, including more than 100 U.S. and international stakeholders and subject matter experts from academia, government, commercial maritime, and Silicon Valley. (The listing of participants is provided in the report section, "Workshop Participants.") Dedicated collaboration sessions with 19 pre-assigned teams were established to identify a problem in the maritime domain that could be solved using advanced data science in the format of a data competition. A full recommendation list is provided in the "Recommendations" report section. Examples of recommendations include:

- Dark Target Data Fusion using open source and commercial data with the goal of establishing a standard approach (taxonomy) to identify dark targets and vessels of interest.
- Improved anomaly detection for illicit cargo with the goal of producing a prioritization of container selection for interdiction based on historical shipping and interdiction data.
- A predictive system to combat piracy in partnership with an insurance company.
- Identifying patterns of fishing activity within certain zones, with the goals of optimizing maritime patrols and reducing IUU fishing.

The GMF workshop inspired participants with new and innovative data platforms and analytics yielding actionable recommendations to make sense of existing maritime data, in order to advance strategic, operational, and tactical decision making. The workshop also fostered new relationships and collaboration much needed to tackle the complexity of maritime domain awareness, while strengthening our existing partnerships with the United Kingdom, Canada, and Australia.

Welcome Remarks

Dr. Eugene Tu

Remarks by Dr. Eugene Tu, NASA Ames Research Center Director, opened the workshop. As NASA's center in Silicon Valley, California, the Ames Research Center contributes to virtually every major NASA mission and initiative via expertise in the following core areas: entry systems, supercomputing, NextGen air transportation, airborne science, low-cost missions, biology and astrobiology, exoplanets, autonomy and robotics, lunar science, human systems integration, and wind tunnels. Dr. Tu described how Ames provides an integrated environment including world-class high performance computing (HPC) resources and services customized to meet NASA's unique needs for Earth and space science, space exploration, and aeronautics, serving users across the country from NASA centers, academia, and industry. He also highlighted how the NASA Advanced Supercomputing (NAS) facility at Ames continually expands its supercomputing and storage resources, including Pleiades, one of the fastest computers in the world. NAS also provides comprehensive user services—application support, large-scale data analysis and visualization, network support, and user environment customization—to broadly accelerate NASA's science and engineering activities, and thus to enhance the success of NASA's mission.

Rear Admiral
Elizabeth Train

RADM Train, Director of the NMIO, and Commander of the Office of Naval Intelligence, welcomed participants and shared her experience, in 1984, at the old Naval Air Station Moffett Field as an air intelligence officer for Patrol Squadron (VP) 19, providing intelligence support for P-3 maritime operations. While some of the issues are common to her experience as an Ensign supporting air crews executing maritime missions, she highlighted the opportunity we now have with advanced computing and data analytic capabilities to make sense of the data.

RADM Train introduced the GMF, set in Silicon Valley, as specifically building upon outcomes from the immensely successful GMF held in Venice, Italy in May 2014 – on "Maritime Data Acquisition and Sharing Technologies, International Opportunities to Enhance Maritime Domain Awareness" or MDA. NMIO was established to coordinate maritime intelligence integration and maritime domain awareness among federal partners and across the GMCOI, which includes state and local maritime agencies, international maritime partners, maritime industry, and maritime academic and R&D partners. RADM Train emphasized that the purpose of this GMF, entitled "Turning the Corner in the Maritime Domain - Leveraging Data to Achieve Effective Understanding," is to improve understanding and to ultimately promote more effective decision making.

Dr. Eugene Tu (left), and Rear Admiral Elizabeth Train

RADM Train highlighted the potential threats our individual nations face from the convergence of "For Profit" illicit activity and transnational crime in the maritime domain, such as human smuggling and trafficking, drug and weapons smuggling, and illegal unregulated and unreported fishing. Knowing that smugglers, traffickers, criminal, and terror organizations, among others, seek to exploit seams within the maritime domain, she asked, "How can we analyze the data better to find them and put them out of business, and how can we leverage the data to improve our mutual maritime domain awareness and really achieve effective understanding?" RADM Train challenged participants to ask these kinds of questions throughout the forum, thinking not just about about the here and now, but also asking what future data analytic capabilities will help us make sense of the tremendously diverse maritime domain, so we can ensure our mutual maritime security and mitigate emerging threats 10 to 25 years from now. Lastly, RADM Train invited everyone to actively participate in the panel discussions and the sessions that followed.

Keynote (Day 1)

Dr. Eng Lim Goh

The Day 1 keynote by Dr. Eng Lim Goh, Vice President and Chief Technology Officer at Silicon Graphics International Corp. (SGI), discussed the technology and implications of High Performance Computing (HPC) and Data Intensive Computing (DIC). HPC or supercomputers are used for a wide range of computationally-intensive tasks in various fields, including quantum mechanics, weather forecasting, climate research, oil and gas exploration, and molecular modeling. DIC is a class of parallel computing applications that use a data parallel approach to process large volumes of data, typically terabytes or petabytes in size, which are typically referred to as "big data."

Dr. Goh presented examples of HPC with the data simulations from NOAA and NASA for detecting earthquake ground motion, tsunamis, and tornado touchdown. Other commercial examples of HPC include simulations for aircraft design applications, oil exploration, and designing the high-technology swimwear fabric (LZR) used in the Racer Suit manufactured by Speedo. Another example of applying HPC to solve a complex problem is optimizing medical scans to reduce the amount of radiation and still improve the image. Dr. Goh characterized HPC as typically outputting (generating) large volumes of data from a smaller stream of data input.

Moving to DIC, Dr. Goh described applications for DIC where massive data is inputted producing (synthesized) data output. He provided examples of such DIC to include NASA's search for exoplanets with the confirmation of 1,883 planets and 1 planet with Earth-like characteristics, tracking infections by the locations of patients through cellular phone data, the scanning of postage to reduce postage fraud, and genomic research to increase wheat yield. Dr. Goh stressed the importance of not filtering data in DIC.

Dr. Goh also discussed various techniques for effective analytics using HPC and DIC with the important role of asking the right question or objective, in order to guide machine learning and serendipity or intuition to observe seemingly unrelated disparate information. He provided the example of the accidental discovery of microwave oven technology, when an engineer noticed that microwaves from an active radar set he was working on started to melt a candy bar in his pocket. Lastly, coupling the output of HPC (usually visualized) to DIC (sensemaking) may be a viable strategy for advanced analytics.

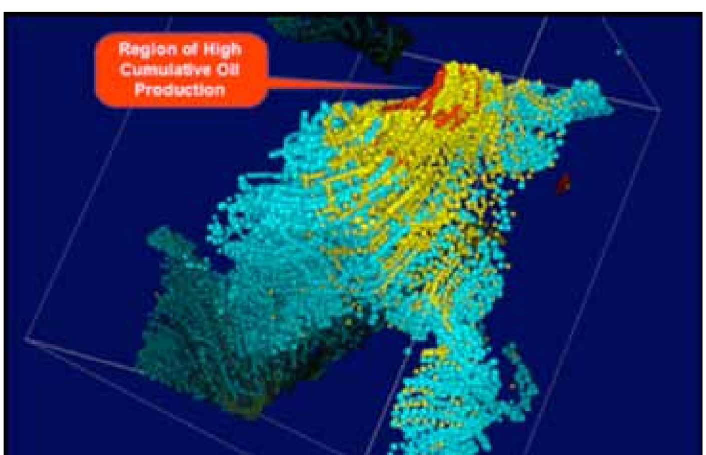
Session I: Platforms and Capabilities

Left to Right – Mr. Pellerin, Dr. Mehotra, Mr. Sternfeld, Mr. Langlois

Mr. David Pellerin

Mr. David Pellerin of Amazon Web Services (AWS) presented use-cases and trends of HPC in the AWS Cloud. The benefits of cloud computing include immediate globally-accessible computing infrastructure and the elastic nature of adding or removing capacity. The anatomy of a cloud platform starts with an infrastructure foundation including network, compute, and storage services; administration and security services; data services; and application services. Examples of using AWS as a platform for applications include Esri's Geographical Information System (GIS) and Stochastic Simulation.

Mr. Pellerin addressed the question of cloud computing security, citing the biggest and most conservative customers meeting their security requirements with a better security profile than what they can deliver internally. The AWS cloud infrastructure has been designed and managed in alignment with regulations, standards, and best-practices, including HIPAA and ISO 27001.



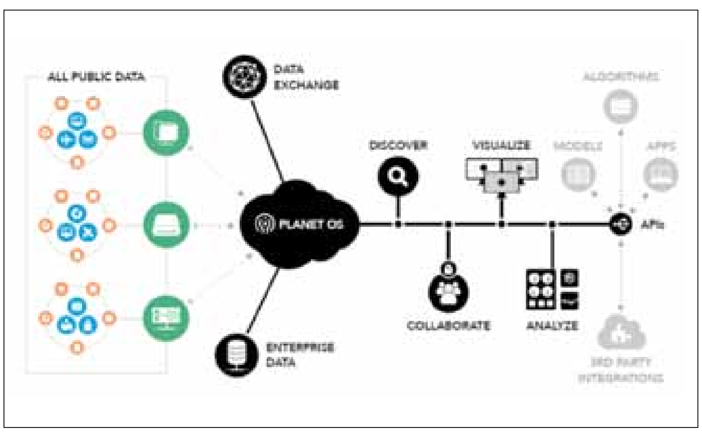
Cluster Analysis in 3-Dimensions
(http://stochasticsimulation.com/resassure)

An example of high-throughput computing is the Large Hadron Collider (LHC) at CERN that includes over 6,000 researchers from more than 40 countries and produces approximately 25 petabytes (1000^5 bytes) of data each year. The ATLAS and CMS experiments use AWS for Monte Carlo simulations and analysis of LHC data. An example of DIC provided by Mr. Pellerin was the Square Kilometer Array (SKA) that will link 250,000 radio telescopes together, creating the world's most sensitive telescope. The SKA will generate zettabytes (1 ZB = 1000^7 bytes = 1000 exabytes = 1 billion terabytes = 1 trillion gigabytes) of raw data, publishing exabytes annually over 30 to 40 years. In life sciences, the Baylor Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) project performed genomics analysis on 14,000 participants generating 24 terabases of sequencer content each month, 1 petabyte of raw data storage, and 21,000 AWS compute cores at peak. Initial analysis was completed in 10 days. Other applications included molecular dynamics simulation and the Financial Industry Regulatory Authority's (FINRA) financial regulation analysis of billions of daily market events. An example of global collaboration for global manufacturing is General Electric's (GE) Crowd-driven Ecosystem for Evolutionary Design (CEED) manufacturing platform. GE uses AWS to connect people, materials, models, simulation, and equipment in an ITAR-compliant, secure, and distributed global environment.

AWS has three consumption models: On-Demand, Reserved, and Spot. In On-Demand consumption, compute capacity is paid by the hour with no long-term commitment and is suited for spiky workloads or to define needs. Reserved consumption is for committed utilization at a significant discount on the hourly charge. For time-insensitive or transient workloads, unused capacity can be consumed at a Spot Price based on supply and demand. Lastly, Amazon has a fully managed service for Machine Learning (ML) to develop, train, and deploy predictive models based on its ML experience in supply chain management, fraudulent transaction identification, image classification and analysis, and catalog organization. Further resources for AWS can be found at aws.amazon.com/hpc and aws.amazon.com/big-data/.

Mr. Rainer Sternfeld, and
Dr. Christopher Clarke (on-screen)

Mr. Rainer Sternfeld and Dr. Chris Clark presented the Planet OS platform with the theme of "What on Earth Do We Know?" The Planet OS platform started with the aggregation of ocean data and is expanding to include all sensor data. The cloud platform for industrial sensor networks provides sensor data discovery and exchange via search and exploration, visualization and analytics, data management, marketplace, and Application Program Interfaces (APIs). The goal is providing one system to access multiple data sets through a single interface that indexes the world: i.e. a cloud platform for industrial sensor networks with no single organization owning or holding the data thereby allowing for discovery and accessibility. Mr. Sternfeld highlighted how finding and accessing the right data is very hard and how this system makes it easier.



Planet OS is a part of NOAA's Big Data Project bringing NOAA's data to the cloud with AWS. Tens of thousands of devices are deployed in the ocean, on land, and in space with hundreds of scattered web services yielding 26,595 NOAA datasets with ISO-19139 metadata and 4,894 NOAA datasets with OpenDAP interface. The project challenge is managing the multitude of data sets, growing by 10 TB per day, and addressing criticality, storing, processing, dissemination, indexing, uniformity, and interoperability.

The case study presented by Dr. Clark provided insight into the sound field analysis performed for the U.S. Navy ensuring environmental compliance of sonar use, thereby protecting marine mammals from harm. The challenge involves merging data sets from 28 C4I sensors, overlaying ship tracks to visualize the tracks, and then seeing the ships' effects act on the oceans' noise levels.

The biggest deployment of Planet OS is Marinexplore.org with 35 organizations, 43,000 data streams, and 8,000 users. The variety of devices and data types (oceanic) include acoustic detections, observer sightings, satellite tagging, imagery analysis, aerial monitoring, vessel Automatic Identification System (AIS), data acoustic recordings, acoustic models, buoys and floats, and wave gliders.

Mr. Justin Langston

Mr. Justin Langston, of the firm Skybox (recently acquired by Google), discussed how the Skybox platform is monitoring important economic locations and providing "good enough" images in real time. The Skybox platform provides access to sub-meter satellite imagery and high-definition video. The application of its collection and data analytic services include monitoring: crop health and forecasting crop yields, refugee movements and infrastructure development in conflict areas to aid humanitarian efforts, high value assets for change and informing risk exposure models to increase efficiency and profitability of insurance models, oil storage containers with sub-meter imagery for changes in volumes to inform commodity trading decisions, and ships entering and exiting ports with HD video to inform supply chain optimization decisions, validate AIS data, and analyze container activity in ports. The SkyNode is a compact system that gives access to directly task and downlink from the Skybox constellation.

The challenge targeted by Skybox is to understand globalization and interdependence of activities of opaque markets and economic bases. Mr. Langston described the Skybox approach of using technology that already exists, such as small satellites and existing sensors, to provide GIS data that is distilled down for usability and decision making. While the service has not been commercially released, Skybox is currently testing it with trusted customers.

Session II: Data Mining and Data Analytics

Dr. Paul Shapiro

Dr. Paul Shapiro of NMIO moderated Session II on data mining data analytics and introduced the panel with a discussion on complex adaptive systems. Understanding maritime activity involves many disciplines with the maritime domain at the intersection of geo-political dynamics, the world economy, and the environment, from weather to the climate. Each of these systems is in of themselves complex. Dr. Shapiro stressed the ultimate goal of extracting simple rules from complex behavior.

Mr. Alan Broder

Mr. Alan Broder of Novetta Solutions presented entity analytics in the maritime domain. He identified the problem of analysts and other end-users currently finding themselves spending 90% of their time organizing and searching for information and only 10% analyzing, with many organizations employing teams of analysts simply for the process of searching. The current approach is neither efficient nor cost effective, and with data only growing, that problem will only worsen.

Mr. Broder described three types of data: structured, semi-structured, and unstructured. While unstructured and semi-structured data are the most difficult to get at, internally and externally, the greatest insights are gleaned from all three data types working together. The question for analysts is, "How do you link these three sources together in a sound, repeatable way that can deal with dirty and fragmented data?"

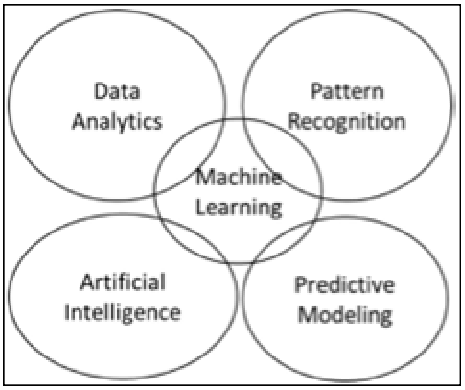
There is no single way to manage and organize this data from an enterprise perspective. The entity analytics technique provides a unified view across systems and sources to identify actors, assets, relationships, and activities hidden in the data by automatically grouping data into logical entities such as people, organizations, places, ships, and other assets. Entities are discovered by massively and fuzzily analyzing combinations of attributes.

Mr. Broder stressed that data consumers should care about all associated data, and not just ship tracks, in order to understand and derive advantage. Unstructured and semi-structured data is prevalent in the maritime domain. If all the dots are not connected, the insights are less trustworthy. Entity resolution can resolve billions of records within hours and link structured with unstructured data, uncovering new connections. By building foundational entity indices of people, organizations, locations, product, and events, organizations have the ability to understand how a person or product is represented in different systems across the enterprise.

Dr. Sangram Ganguly of the NASA Ames Research Center's Earth Science Division discussed machine learning applications for earth sciences with a focus on satellite image classification. In data-intensive scientific research, the Big Data conundrum consists of storing more data, accessing large quantities of data faster, getting a better understanding of what the data tells us (structured vs. unstructured), integrating Software as a Service (SaaS), cloud computing solutions, and data access efficiently and following industry standards. Dr. Ganguly recommended reading The Fourth Paradigm for further insight into data-intensive scientific discovery. Drawing from the "American Geophysical Union Sessions IN006. Big Data in the Geosciences: New Analytics Methods and Parallel Algorithms," Dr. Ganguly outlined solutions to the big data conundrum from accessing and sorting data in an efficient manner using open database solutions and frameworks like Mongo DB and Apache Hadoop, to writing massive parallel applications and distributed jobs, and deploying cloud computing solutions such as Openstack and AWS.

Dr. Sangram Ganguly

Dr. Ganguly shared NASA's vision through the NASA Earth Exchange (NEX) to provide "Science as A Service" to the Earth science community addressing global environmental challenges. 'Science As a Service" refers to the NEX ready-to-use data, models, tools, and workflows via a portal, a development sandbox, and HPC. The classes of NEX Big Data projects range from fully distributed data processing with no inter-process dependencies (data sizes: 100TB to 5PB) to data-mining with some inter-process data dependencies (data sizes: 300TB to 2PB) to analytics and science applications (databased query systems: 1 to 10 TB).

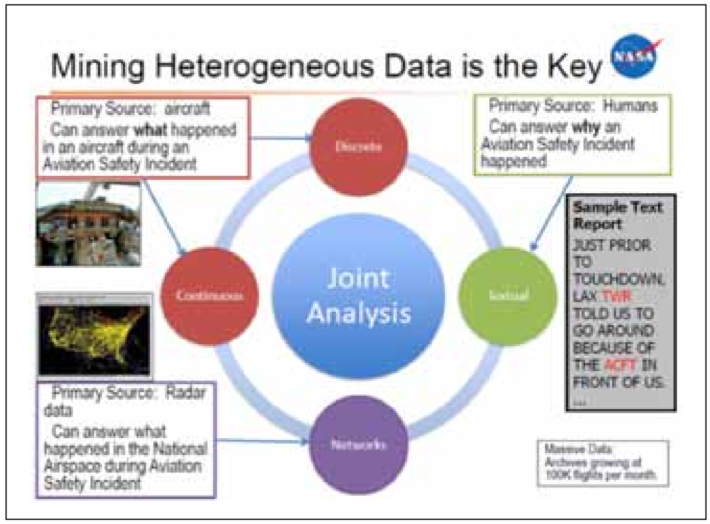


Dr. Ganguly defined machine learning as a science that enables us to teach computers to take actions without being explicitly programmed to do so. Central to analyzing big data, machine learning is now everywhere from Google's search to Facebook's face recognition to Apple's Siri. Types of learning include supervised, reinforcement, unsupervised, and semi-supervised. Supervised learning determines a function that maps inputs to outputs using labeled training examples while unsupervised learning discovers latent patterns in data without supervision. Reinforcement learning performs a goal in a dynamic and volatile environment without supervision such as driving a vehicle or playing a game against an opponent. Lastly, semi-supervised learning lies between supervised and unsupervised learning with some of the training data being unlabeled. The NEX projects showcasing machine learning applications included: Satellite Anomaly Workflow, Global Drought Monitoring, Web Enabled Landsat Data (WELD) Processing, North American Forest Dynamics (NAFD) Processing, Carbon Monitoring System (CMS) Processing, Supporting National Climate Assessment (NCA), and Agricultural Monitoring. Further information on NEX is available at https://nex.nasa.gov.


Dr. Rodney Martin

Dr. Rodney Martin of the NASA Ames Research Center Data Sciences Group presented "Data Mining, Machine Learning, and Big Data at NASA and Beyond." He described the types of problems tackled by NASA, an example of a specific problem, where data mining fits in, why "Big Data" is a misleading phrase, and how NASA releases its results. The research and development (R&D) of methods in data mining, machine learning, and knowledge discovery is performed in close collaboration with domain experts and tool developers to help advance NASA missions by turning data into useful insights. Problems areas span aeronautics, earth science, space science, space science, and aviation safety.

The specific example provided by Dr. Martin delved into anomaly detection in aviation. Data driven methods are used to discover anomalies by learning statistical properties of the data



and finding which data points do not fit. Mining heterogeneous data is the key to analyzing discrete and continuous sources of data from the aircraft during an Aviation Safety Incident, along with textual information from human operators and related data such National Airspace radar data. Archives of aviation data are growing by 100,000 flights per month. Anomalies identified include drop in airspeed during takeoff, possible mode confusion, and unstable approach.

Dr. Martin's discussion on the term "Big Data" focused on volume, variety, velocity, and veracity. While most people think of volume only, volume is often the least important part. The variety of the data may encompass numeric, text, models, and networks at various resolutions and accuracies. Velocity refers to the speed at which new data is generated and moved around. Lastly, veracity refers to the varying levels of noise, uncertainty, bias, or processing errors.

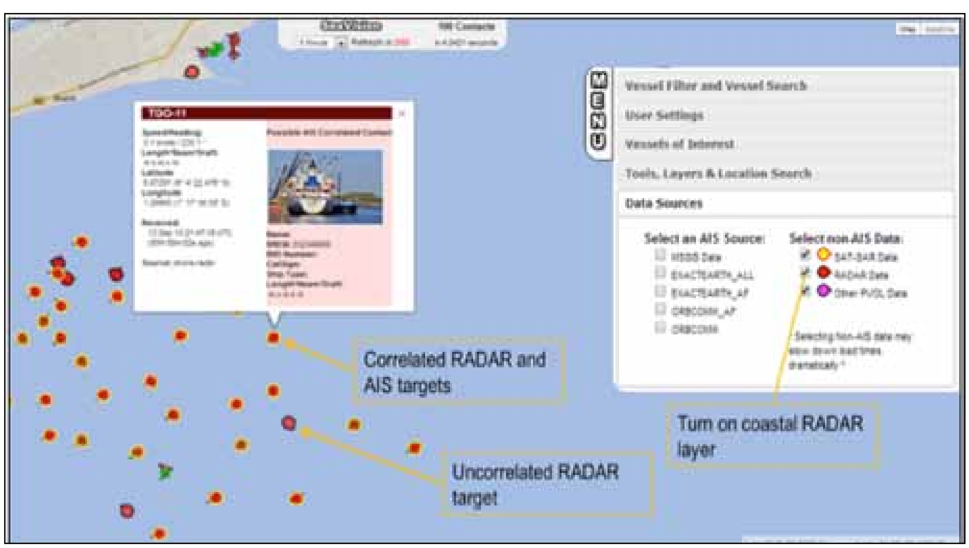
Session III: Data Visualization and Accessibility

Left to Right – Dr. Everton, Mr. Stastny, Dr. Henze, Dr. Vera

Mr. John Stastny

Mr. John Stastny, Space and Naval Warfare Systems Command (SPAWAR) Pacific, spoke on the topic of "Maritime Data and Analysis Visualization for Improved Maritime Safety and Security." He identified the maritime domain's unique environment and many challenges, to include a wide variety of objects (small canoes/rafts, fishing vessels, large tankers/cargo vessels), each of which presents unique challenges with respect to MDA (detection and tracking), with the ocean as a vast, mostly unoccupied space in which it is difficult to detect and track all components continuously, and surveillance methods are susceptible to "high clutter." There are mostly well-behaved vessels with only a few "bad actors." Simply viewing all activity will overwhelm any operator, and in the face of these challenges, piracy, sea robbery, smuggling of narcotics, and illegal fishing threaten national security and international commerce.

Mr. Stastny provided insight into the Cooperative Interagency Partnership for Maritime Domain Awareness (CIOP-MDA). The project aims to improve the web-based Common Operating Picture (COP), facilitate information sharing between Maritime Operational Centers (MOCs), identify and address gaps in operational maritime capability, and build MDA capacity in partner nations. The COP, named SeaVision, and used by the United States and partner nations, is capable of displaying on top of a base map terrestrial AIS, along with satellite AIS, coastal radar, satellite imagery, and IHS Fairplay static data. The automated correlation of satellite imagery with terrestrial and satellite AIS quickly identifies non-AIS-emitting vessels to assist in the planning of operations and patrols. Using industry standard data formats output from most commercial radar processing units, SPAWAR worked with partners to integrate coastal radar vessel tracks into the COP and developed and tested a correlator to fuse coastal radar with other track data. SeaVision allows both regional sharing of coastal radar data and the ability to quickly identify vessels not emitting AIS, thereby providing key information about the locations of unknown vessels, i.e., possible IUU fishing vessels.



Mr. Stastny provided a specific example of visualization with the automated U.S. Coast Guard Port State Control Targeting Prioritization score calculation. The system visually presents a boarding priority for each vessel on the operator display (Red: Priority I, Yellow: Priority II, Green: Not a Priority) with an explanation of the scoring and classification. The calculation is performed for approximately 70,000 vessels and updated twice daily, based on the Port State Control Safety and Environmental Protection Compliance Targeting Matrix.

Dr. Chris Henze

Dr. Chris Henze of the NASA Advanced Supercomputing Division presented on data management and visualization. He discussed working with NASA's largest super-computer, Pleiades, which consists of 162 SGI racks, 11,280 nodes, interconnected by 65 to 70 miles of wires. A total of 211,360 CPU cores and 723 terabtyes of memory are capable of computing 5.33 Pflops (a Pflop equals one quadrillion Floating-Point Operations Per Second).

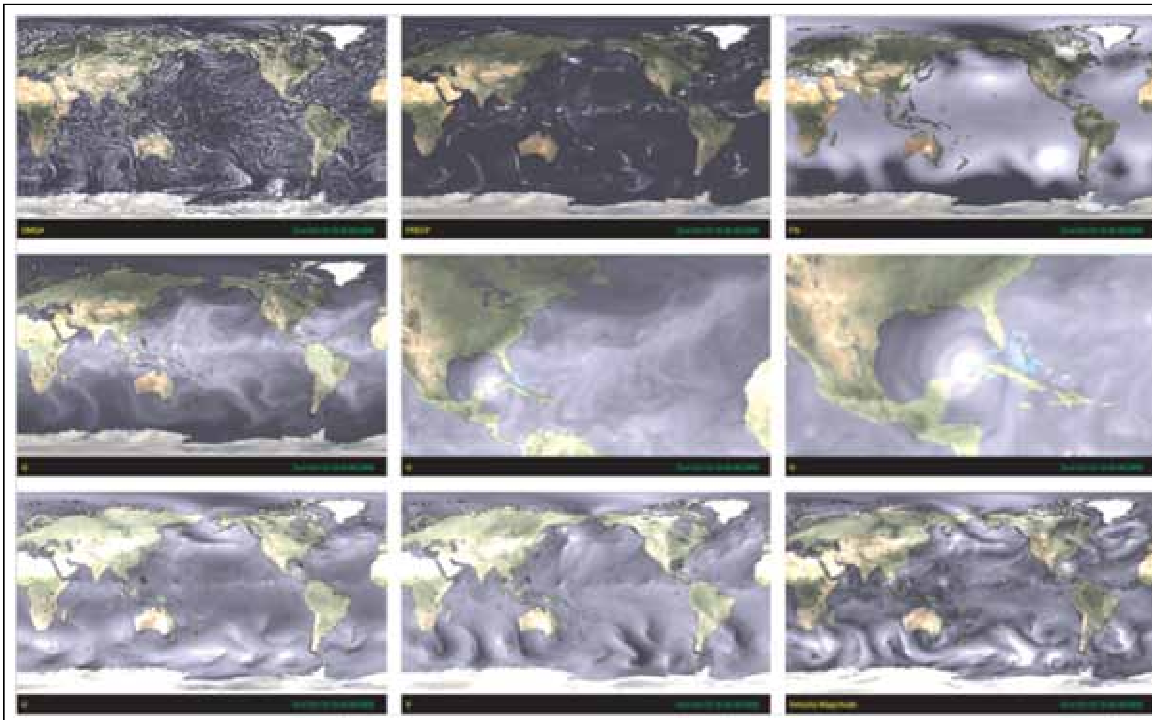
Dr. Henze emphasized the inordinate amount of data management effort required to effectively use parallel computing by splitting the domain over the processors. The technique of domain decomposition involves dividing a computation into a "local" part, which may be done without inter-processor communication, and a part that involves communication between processors.

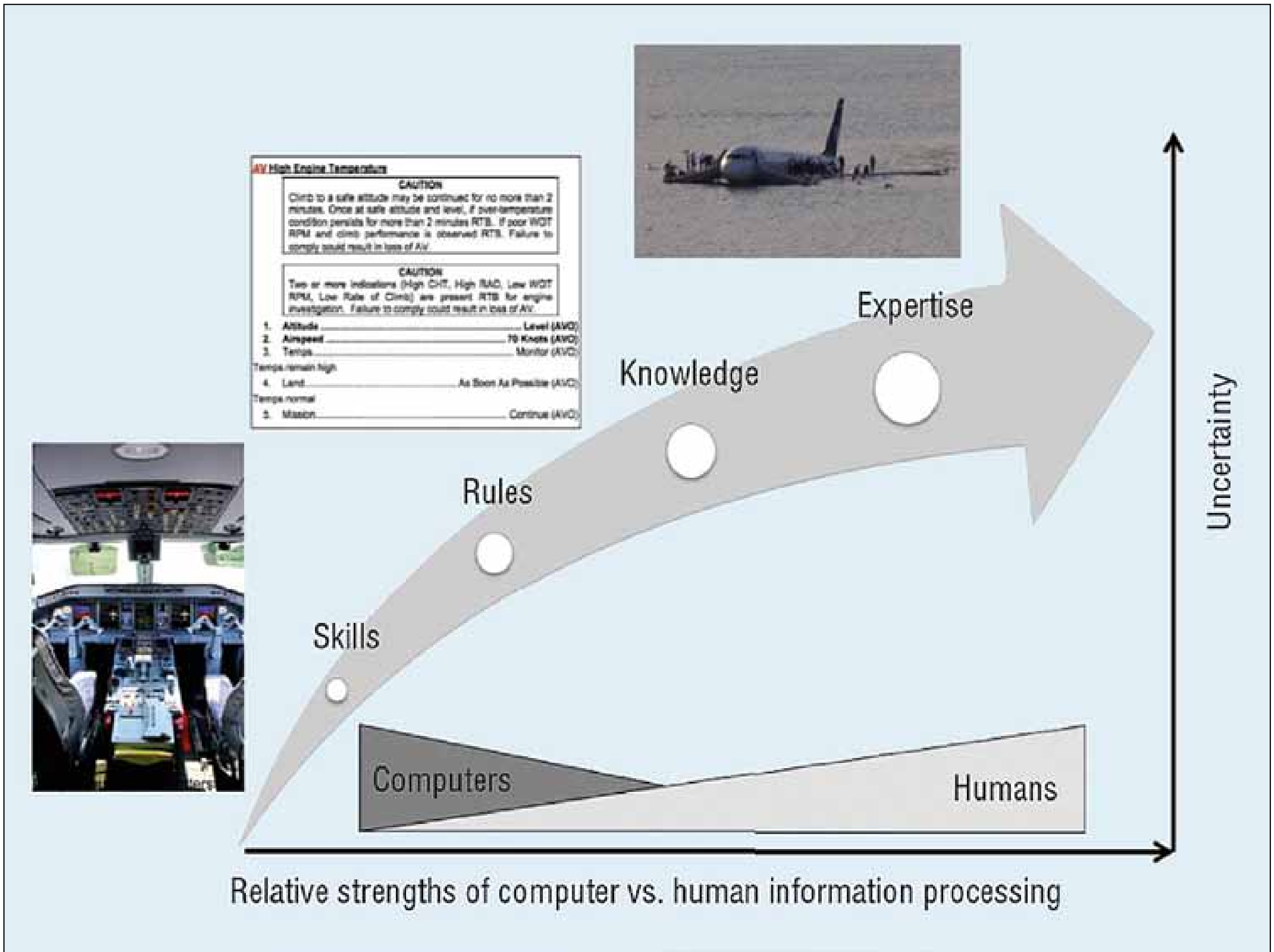
Dr. Henze also presented the challenge of visualizing data with examples of simulations, including a rocket motor, Ospry blade flow, and a global ocean model.

Dr. Alonzo Vera

Dr. Alonzo Vera, Chief of the NASA Human Systems Integration Division, presented the topic of "Adapting Visualization to Human Perception and Expertise." Beginning the presentation with machine intelligence, Dr. Vera identified four areas of related development including: Big Data, Deep Learning, networked learning and cyber-physical systems, and Moore's Law (bigger and faster computers driving change with increasing velocity). In considering human interaction, "The Autonomy Paradox" (Blackhurst, Gresham & Stone, 2011) describes how autonomy does not get rid of humans, it changes their role. Citing the example of the Littoral Combat Ship (LCS) built to be operated by 45 sailors, investments in automation have led to higher costs because more humans are needed to deal with the complexity of newer systems. The LCS turned out to need 65 sailors for operation at much high than expected qualifications. The take-away from the LCS example is the need to concurrently design for autonomy and humans.

Dr. Vera discussed the human cognitive architecture as being neither *tabula rasa* (Latin often translated into English as "blank


(Image of ocean models as depicted in the article "Concurrent Visualization in a Production Supercomputing Environment" by David Ellsworth, Bryan Green, Chris Henze, Patrick Moran, and Timothy Sandstrom published in IEEE Transactions on Visualization and Computer Graphics, vol. 12, no. 5, September/October 2006.)



From Cummings, M.L., “Man vs. Machine or Man + Machine?” IEEE Intelligent Systems, (2014) 29(5), p. 62–69.

slate”) nor randomly constrained. The key characteristics of human problem solving focus on induction (moving from specific observations to broader generalizations) rather than deduction (general to the specific) questioning “Why did this happen?” and “Can this be done better?,” and using heuristics and biases. Dr. Vera discussed how “small data” needs to be treated differently from “big data” where human insight and intuition work well, and the need to work through “peripheral” channels (through which we easily receive data without being particularly focused on it).

Dr. Vera also discussed the teaming of human and machine intelligence with the goal of designing the human into the process with a focus on the interface. He challenged participants to consider with respect to Big Data, “Not how to store or how to query, but rather how do you impose human intuition on data of this scale?”

Keynote (Day 2)

Mr. Kshemendra Paul

Mr. Kshemendra Paul, PM-ISE, delivered the Day 2 keynote address on challenges and efforts underway to achieve the vision of national security through responsible information sharing.

After the terrorist attacks of 9/11/2001, the Intelligence Reform and Terrorism Prevention Act of 2004 established the position of Program Manager to "plan for and oversee the implementation of, and manage the ISE," and to be "responsible for information sharing across the Federal Government." The PM-ISE serves as a change agent and center for innovation and discovery in providing ideas, tools, and resources to mission partners, who then apply them to their own agencies or communities.

Mr. Paul defined responsible information sharing as providing the right information to the right people in the right form at the right time to enable them to do their jobs better, faster, cheaper, and easier. This applies to both openly available data and to more sensitive data, which must have a level of safeguarding. PM-ISE focuses on information sharing to foster better decision making, to treat information as an asset, and to realize its full value.

Mr. Paul provided insight into the work taking place across a vast decentralized enterprise through national approaches to interoperability; common operating models; consistent, transparent and distributed policies; and integrated capabilities and shared services. The main challenges are evolving and converging threats of transnational crime, cyber-attacks, and violent extremism that are both national and international in nature and often do not respect administrative boundaries.

Mr. Paul discussed the ISE mission portfolios, including statewide and regional ISE, watchlisting, screening, encounters, transnational organized crime, cybersecurity information sharing, domain awareness, and incident management. He provided examples of each portfolio that has aspects that can be improved through technology such as:

- analyzing crime concentration patterns to optimize deployment of limited resources for state and regional place-based policing,
- enabling cyber analysis to identify the intersection of physical and virtual threats,
- the need to track travel patterns and times of known and suspected terrorists for effective watchlisting and screening,
- domain awareness (maritime) as the continuous awareness of ships, cargo, and crew movements in time and space,
- situational awareness and coordinated response to natural disasters and man-made incidents, and
- the ability to analyze patterns of crime and movement of criminals.

Mr. Paul emphasized the need to promote strong collaboration between the public and private sectors. PM-ISE enabled this collaboration through several initiatives, including the Standards Coordinating Council, Project Interoperability, and the Center for Collaborative Systems for Security, Safety and Regional Resilience (CoSSAR). PM-ISE established a Standards Coordinating Council composed of government agencies, technology industry associations, and standards development organizations to work across organizational lines to collaboratively identify, develop, and promote information sharing best practices. More information about the Council is available at http://www.standardscoordination.org/. Project Interoperability employs a combination of management practices and technical approaches to enable successful information sharing (http://project-interoperability.github.io/). Mr. Paul also highlighted two documents demonstrating a whole-of-government approach: "The National Maritime Domain Awareness Architecture Plan" and "Consolidated Vessel Information and Security Reporting (CVISR)."

Mr. Paul challenged the workshop participants to "give us some more success stories" for the ISE blog (http://www.ise.gov/) by using and extending the work already going on; taking advantage of the fact PM-ISE already has an underlying framework; and to consider the distributed, decentralized, and coordinated approach.

Session IV: IUU Fishing and the Maritime Environment

Left to right, Dr. Mittleman, Ms. Austin, Dr. Nolker, Mr. Jacobs

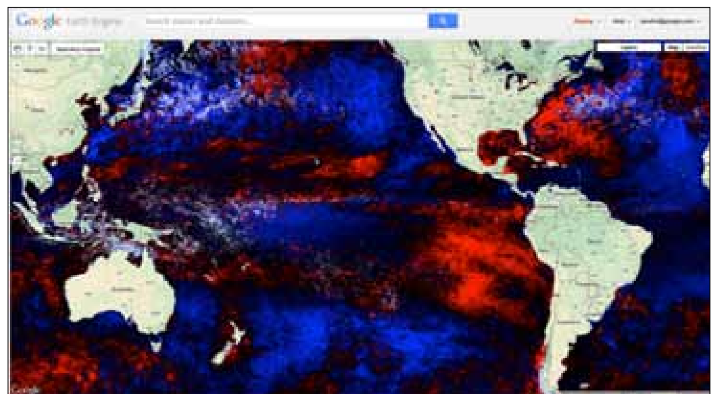
Ms. Jenifer Austin

Ms. Jennifer Austin, the manager of the Google Oceans program, gave a presentation co-written with Mr. Michael Jones, the former Chief Technology Advocate at Google. Google Earth is a virtual globe, map, and geographical information program that was originally named EarthViewer 3D and created by Keyhole, Inc, a Central Intelligence Agency (CIA)-funded company acquired by Google in 2004. [https://en.wikipedia.org/wiki/Google_Earth]

The development of Ocean in Google Earth began in 2007 and launched in early 2009, allowing users to navigate underwater.

Ms. Austin highlighted the importance of policy, regulations, and enforcement on shaping behavior. She cited the example of Iceland's approach to its need to continuously track vessels, using the TrackWell Vessel Monitoring System (VMS). Iceland requires all vessels to use AIS. If a vessel does not report AIS for more than 20 minutes, they dispatch a rescue helicopter assuming a vessel is in trouble. If the vessel does not require assistance, the operator is responsible to pay for the cost of the rescue assistance. Additionally, Iceland uses bar codes to track the source of fish and products. More information on the Trackwell VMS is available at http://www.trackwell.com/maritime/fishing-authorities/vessel-monitoring-system-vms/.

Ms. Austin described various collaborations to visualize data including vessel track data from the small satellite firm, SpaceQuest, and imagery from Skybox. She also highlighted the Google Earth Engine platform for planetary-scale environmental data and analysis. Google Earth Engine provides satellite imagery and trillions of scientific measurements dating back over 40 years with tools for scientists, independent researchers, and nations. Access to Earth Engine is currently available as a limited release to a small group of partners.



https://earthengine.google.org/#intro

Ms. Austin also discussed the technology partnership between Google, Oceana, and SkyTruth to extract and visualize two years of world-wide fishing activity. The resulting product is the interactive web tool, Global Fishing Watch. Currently in the prototype stage, Global Fishing Watch aims to reveal the intensity of fishing effort around the world and enable anyone to visualize the global fishing fleet in space and time. More information can be found at the following web site: http://www.globalfishingwatch.org/.

Lastly, Ms. Austin discussed several initiatives to deploy sensors including: Google's Trekker spherical photo and mapping technology on a Wave Adaptive Modular Vessel (WAM-V), capturing the entire San Francisco shoreline and underwater Street Views for 40 locations from around the world.

Mr. Todd Jacobs

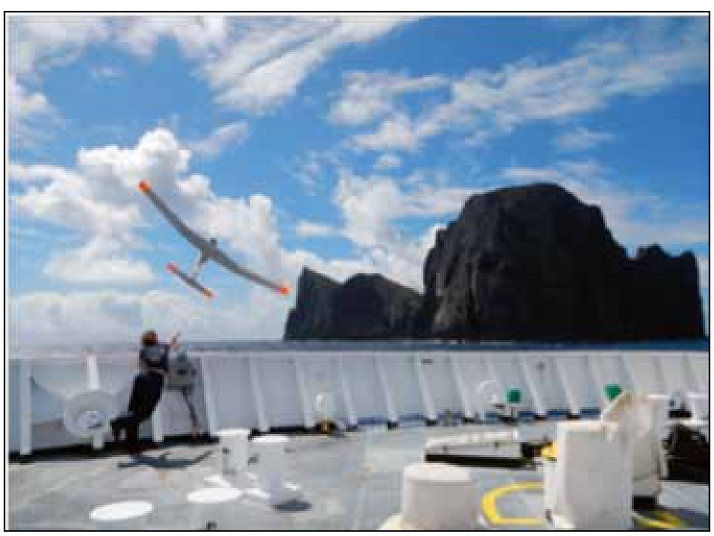
Mr. Todd Jacobs of NOAA provided insight into NOAA experiences with unmanned aerial systems (UAS). NOAA's small UAS history includes the tests of various systems such as ScanEagle in 2007 and 2009, the acquisition of two multi-copters in 2010, and two Puma™ UAS in 2011. ScanEagle is an autonomous unmanned aerial vehicle, developed and built by Boeing and its subsidiary, Insitu Inc., and marketed by ScanEagle® Unmanned Aircraft Systems. The Puma™ AE (All Environment) is a small unmanned aircraft system designed for land-based and maritime operations.

NOAA's UAS are currently in use aboard sea vessels for marine resource monitoring and are being evaluated for enforcement and emergency operations. Mr. Jacobs highlighted the ScanEagle that was on test deployment in Puget Sound in 2007, and the use of ScanEagle aboard NOAA Ship Oscar Dyson in 2009 to study ice seals in the Arctic.

The prerequisites for UAS operations include access to domestic or international airspace; access to well-maintained equipment; availability of trained, certified, and proficient operators and observers; and access to available bandwidth to send from remote operating areas to shore based researchers.

The development of protocols and procedures for UAS usage supports the following missions:

- Living Marine Resource Surveys
- Enforcement and surveillance of fisheries and Marine Protected Areas (MPAs)
- Emergency Response
- Marine Debris
- USCG Arctic Support
- Habitat Mapping
- Enforcement



Mr. Jacobs discussed how data can be leveraged, and offered ideas on technological areas for development to improve the effectiveness of UAS operations. The processing of satellite, hydrophone, and other data would help to identify target times and places to deploy UAS. The analysis of historic flight data in remote oceanic regions to develop statistical likelihood of manned aircraft interaction would help the issuance of FAA Certificates of Waiver or Authorization (COA).

Lastly, Mr. Jacobs provided the following observations on UAS operations:

- the utility of UAS for fisheries enforcement increases with the remoteness of the area;
- vessel-based UAS operations are not that complicated;
- systems that do not require vessel modifications are simpler to integrate and offer much greater potential for "spontaneous" operations against vessels of opportunity;
- UAS will be effective tools for fisheries and MPA surveillance and enforcement once routinely available to be operated beyond line-of-sight (BLOS) in the U.S. National Airspace System (NAS);
- the expense and complexity of operating a Predator class UAS for fisheries and Marine Protected Areas (MPA) surveillance and enforcement are such that operations would likely need to be carried out in conjunction with other compatible missions;
- ScanEagle may be a viable solution for remote ship operations, and;
- airspace is still an issue, at least in the NAS.

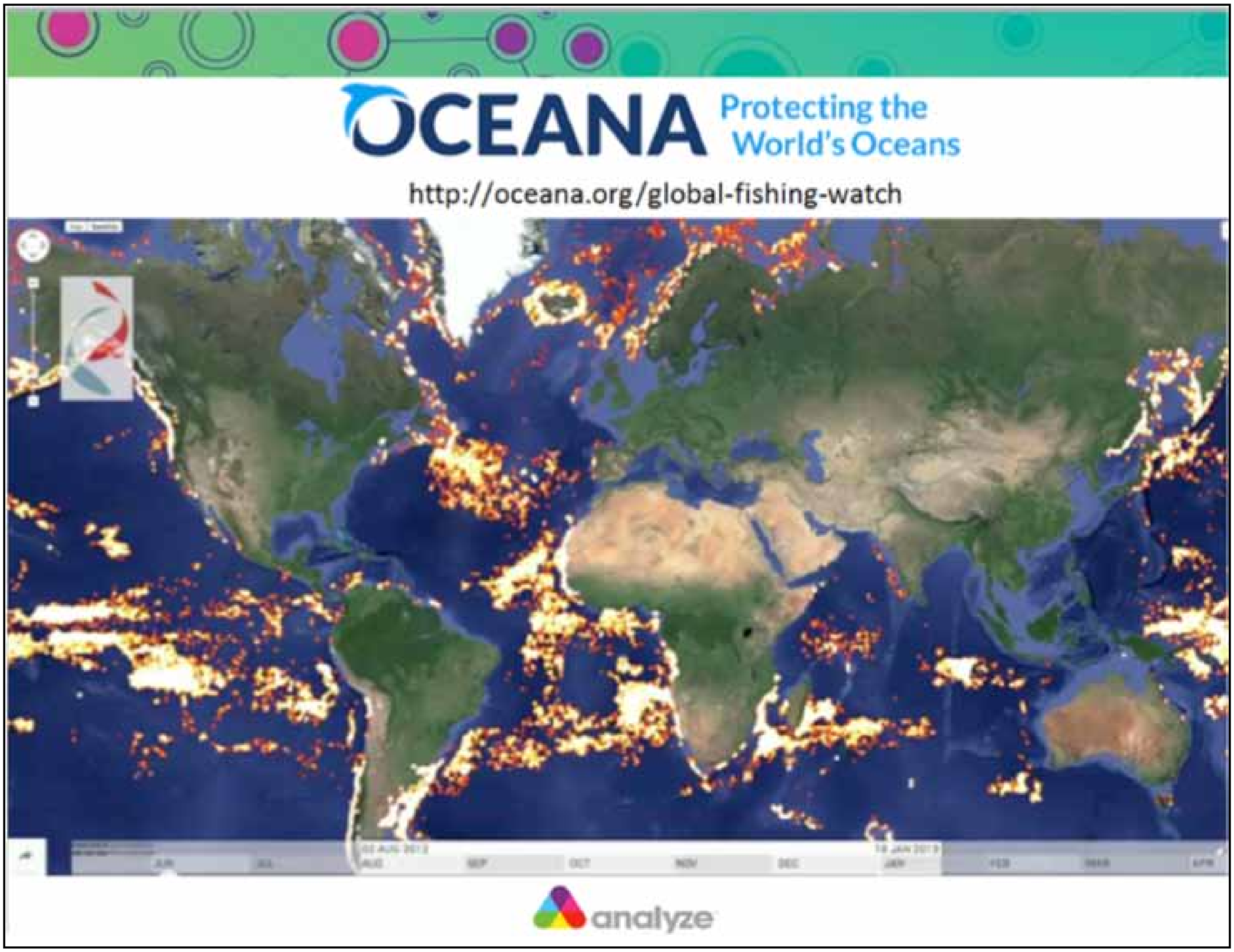
Dr. Robert Nolker

Dr. Robert Nolker of Analyze Corp presented his work in MDA, focusing on motion analytics for behavior profiling. Using SpaceQuest global AIS data, Analyze's behavioral algorithms specifically identified basic fishing patterns based on fine-grained track data and economic incentives. The goal of identifying fishing behavior using satellite AIS data directly confronted the challenge of understanding the characteristics of various types of fishing behavior and activity.

The analytic strategy begins with data acquisition and Extract, Transform and Load (ETL) to clean up the raw AIS sensor data. Next, data analytics and motion analytics employ statistics, machine learning, and visualization to determine how the objects are moving and determine, "What does the data say?" The results of the data and motion analytics inform behavior analytics and determine from observed behavior, a most probable activity, such as fishing, spoofing, or trans-shipment. Consequently, with Dr. Nolker's predictive analytics we can now estimate probability of future actions, such as whether a vessel will continue its current activity, start a new activity, or move to a different location in the future.

The prototype interactive visualization tool of vessel fishing activity was a direct result of the collaboration with SpaceQuest, Analyze, Google, and SkyTruth. Tracks of fishing and other vessels are detected with satellite-based AIS receivers operated by SpaceQuest. Vessel tracks are classified by Analyze to rate the likelihood that a ship was actively engaged in fishing activity at each track point. Each detection is highlighted using a visualization technology developed jointly by Google and SkyTruth. The prototype was introduced to the world by Oceana as Global Fishing Watch.

Currently, in collaboration with University of Southern California (USC), San Diego State University, and the U.S. Coast Guard, Analyze is expanding the research to a Panga Interdiction Predictive model. They are integrating open source data fusion and behavior analytics with the goal of better using law enforcement resources. Other areas of research include the role of satellite cell phones on the open seas and GPS devices in coastal waters. Lastly, Dr. Nolker framed the IUU fishing challenge in terms of finding the "pressure point" to effect positive change through activism, economics, and policy across consumers, businesses, and governments. An "AIS always on" certification to reward the good guys may incentivize positive behavior.



Session V: Interdiction and Safety

Mr. Scott Genovese

Mr. Scott Genovese, Director of the Global Maritime Operational Threat Response (MOTR) Coordination Center (GMCC), introduced the challenges of legally enforcing laws on the oceans. As a jurisdictionally, legally, and operationally unique operating space, the oceans are a complex array of flag, port, and coastal state authorities and maritime zones; an array of ownership, crew, and cargo from potentially dozens of countries on a single vessel; and a vast operating environment with large swaths of ungoverned, under-governed, and ungovernable areas.

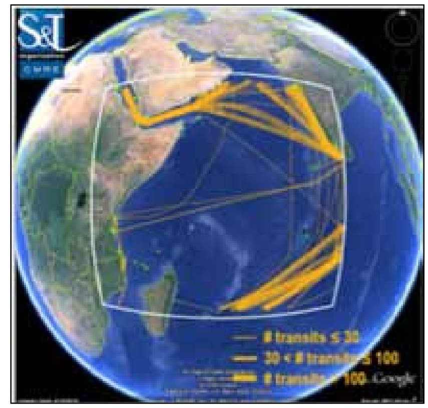
Mr. Genovese highlighted that the "interagency" is a process, not a place. Best practice considerations with whole-of-government information sharing and decision-making must take into account the various cultures, organizational structures, laws, policies, and directives.

Ms. Karna Bryan

Ms. Karna Bryan of NATO's Centre for Maritime Research and Experimentation (CMRE) presented "Data Analytics supported by Big Data Analytics."

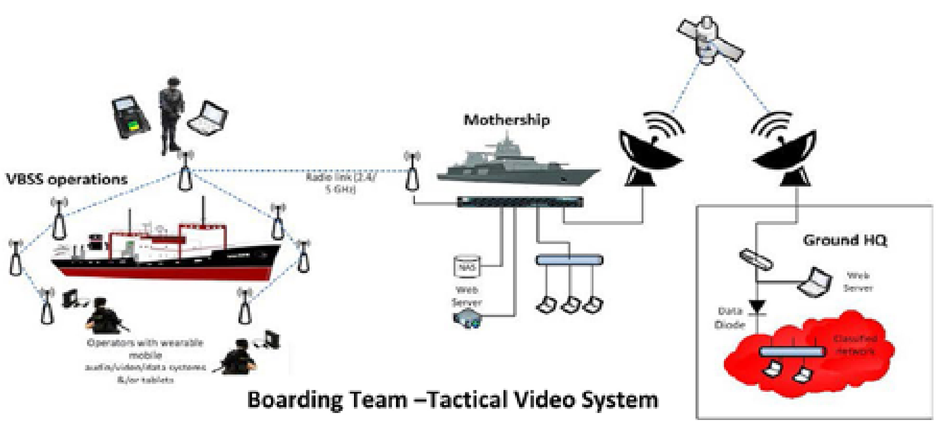
Ms. Bryan defined maritime interdiction as naval operations that aim to interrupt, dissuade, or prevent enemy or illicit activities at sea before they do any harm. NATO's 2011 Alliance Maritime Strategy increased the role of navies in maritime security missions. With the unique capabilities of militaries, analytic capabilities play a central role for Maritime Situational Awareness (MSA) and Intelligence Surveillance Reconnaissance (ISR). Military analysts need to know what's normal so that they can spot what's abnormal.

The Traffic Route Extraction and Anomaly Detection (TREAD) methodology was developed for different levels of intermittency (i.e., sensor coverage and performance), persistence (i.e., time lag between subsequent observations) and data sources (i.e., ground-based and space-based receivers). An unsupervised and incremental learning approach to the extraction of maritime movement patterns converts raw data to information supporting decisions. This is a basis for automatically detecting anomalies and projecting current trajectories and patterns into the future. The aid of automatic processing to synthesize the behaviors of interest in a clear and effective way is required due to the amount of information, which is increasingly overwhelming to human operators. [Pallotta G., Vespe M., Bryan K. (2013) "Vessel Pattern Knowledge Discovery from AIS Data: a Framework for Anomaly Detection and Route Prediction". Entropy, Big Data Issue 15(6), pp. 2218-2245. ISSN1099-4300)]



Ms. Bryan highlighted barriers to using new technology such as accreditation and acquisition cycles and the criticality of interoperability testing. Using information in operational environments requires scenarios and use case development with shared storyboards to promote collaboration and create common understanding between military, industry and academia. Information architectures from multiple perspectives (i.e., Operational Command, Information Processing, and Autonomous Sensors and Platforms) support human-machine hybrid reasoning and improve understanding and decisions.

According to Ms. Bryan, furture challenges include the complexity of the global supply chain and the variety and intermittency of current data. The "Internet-of-Things" future will provide new options, while distributed sensor networks will provide new data fusion challenges, along with additional technology supporting maritime interdiction operations such as tactical video system for boarding teams.



Mr. Neil Watts spoke on behalf of the United Nation (UN) Panel of Experts established pursuant to Resolution 1874 (2009). Within the topic of interdiction and safety, he specifically focused on the Democratic People's Republic of Korea (DPRK) sanctions and illicit flows. SCR1874 (2009) depends on member states to implement sanctions. There have been 16 maritime vessel violation cases including the Chong Chon Gang, interdicted in 2013, and found to be hiding missile system components, and the ANL Australia cargo interdiction of arms and ammunition. The ANL Australia was an Australian-owned Bahamas-flagged vessel, with an Italian shipper utilizing a false Bill of Lading, whose Iran-bound cargo was seized in United Arab Emirates (UAE). The DPRK employs a broad range of techniques to mask flows of prohibited goods, including misleading Bills of Lading, false labeling, misleading declarations, false compartments etc.

Mr. Neil Watts

Mr. Watts detailed what has been learned about the patterns of sanction evasions in trade of goods to include:

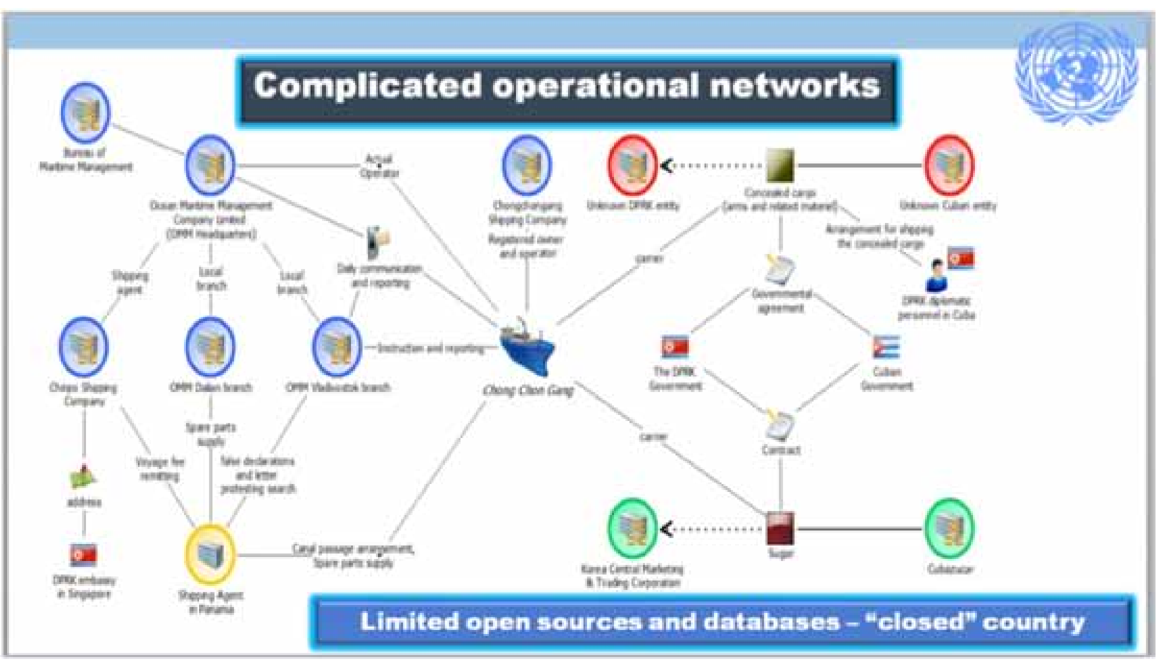
- legitimate trade used as cover;
- multiple layers of intermediaries, shell companies, financial institutions, and small companies;
- multiple countries to conceal origin and recipient;
- falsification of cargo manifest documentation; and
- physical concealment measures of the cargo to deceive cursory physical inspections and trans-shipments, which may include multi-modal transport streams.

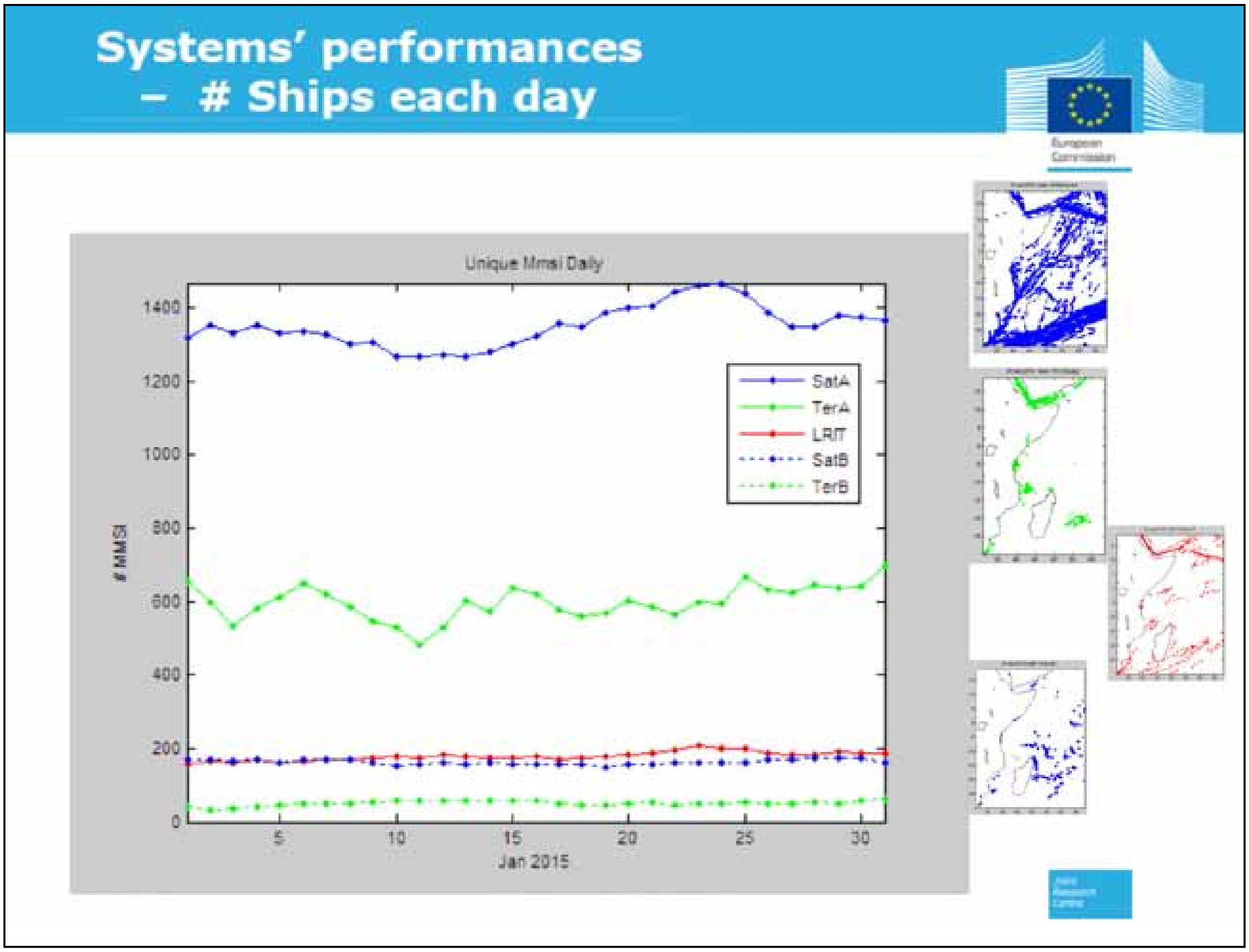
The implication is that legitimate companies, including sea carriers, shippers, shipping agents, and banks, etc., can become caught up in a sanctions violation. The impacts are: ships being delayed while searched, ships diverted for search and seizure, and port disruptions. The challenges include limited resources, lack of data and empirical research, limited access to "closed" countries, vessel tracking, vessel identification, verification, and the identification of the players and entities involved in international shipping industries.

Dr. Harm Greidanus of the European Commission Joint Research Center presented on improving MSA. Ongoing research and development into techniques to improve MSA include: maritime awareness around Africa, performance and quality of AIS and long-range identification and tracking (LRIT), ship route extraction and characterization, use of LRIT data, monitoring fishing activities, the Arctic, radiolocation of AIS signals, and container traffic monitoring.


Dr. Harm Greidanus

Piracy, Maritime Awareness and Risks (PMAR) provides a region-wide MSA currently aimed at building capacity in African countries. Data from up to 17 satellites (Norway, LuxSpace, Orbcomm, exactEarth, SpaceQuest) is provided along with supporting information layers such as predicted wind and wave fields from NOAA's Environmental Research Division's Data Access Program (ERDDA). Projects include PMAR-1 focused





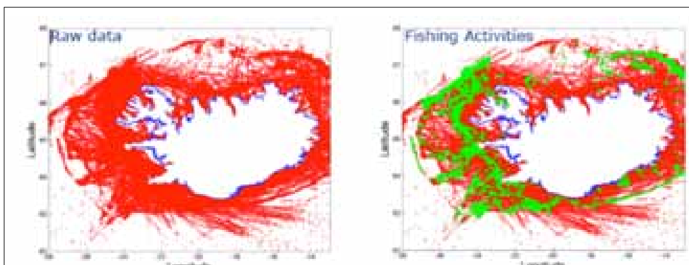
on the Horn of Africa from 2010 to 2012, PMAR-2 in the Gulf of Guinea from 2012 to 2013, and PMAR-MASE from 2014 to 2015 in use by the Indian Ocean Commission, Seychelles, and the Kenya Maritime Authority, Mombasa.

Dr. Greidanus discussed the performance and quality of AIS and LRIT in tracking of ship positions. For Class A vessels (ships greater than 300 gross tons), using satellite AIS, coast AIS, and LRIT data from February 2015, he illustrated ship density maps and statistics per country EEZ and per flag. The system's performance was then illustrated comparing the number of ships detected each day by data type.

Dr. Greidanus provided an analysis of AIS data quality through the case study of tracking the Maersk Tigris for 30 days up to 7 May. With the graphical mismatches in longitude and latitude of the vessel track as reported through the Maritime Safety and Security Information System (MSSIS) system, he clearly demonstrated AIS data errors and the need to "clean" AIS data. Dr. Greidanus presented several case studies, including the impact of piracy on maritime routes in the Western Indian Ocean as seen by LRIT and deriving fishing behavior from motion.

With regard to monitoring fishing practices and sustainability, Dr. Greidanus also discussed the use of genomics and traceability and various EU regulations for food labeling, traceability, and IUU catch certificates. Projects include:

- FishPopTrace: Genetic origin assignment of cod, hake, common sole, and herring;
- FishTrace: Genetic reference catalogue for species identification of all major commercial fish species in Europe (200 species characterized);
- AnchovyID: Genetic markers for anchovy product identification;
- SturSNiP: Genetic markers for sturgeon product (caviar) identification; and
- AqiuaGen: Project to develop tools for the genetic distinction between wild and farmed marine fish.



(Mazzarella F., Vespe M., Damalas D., Osio G.: 'Discovering Vessel Activities at Sea using AIS Data: Mapping of Fishing Footprints', Proc. 17th Int. Conf. on Information Fusion, 2014)

Lastly, Dr. Greidanus presented the system "ConTraffic" used by EU anti-fraud authorities to perform container traffic analysis. With less than 2% of shipping containers being physically inspected, the system uses data on the itinerary of containers for improving risk assessment. Based on the full itinerary of where the goods come from, statistical analysis and data mining along with visual analytics to detect anomalies and outliers, more than 10 million records per day are analyzed.

Session VI: Global Supply Chain

Mr. Michael Rodriguez

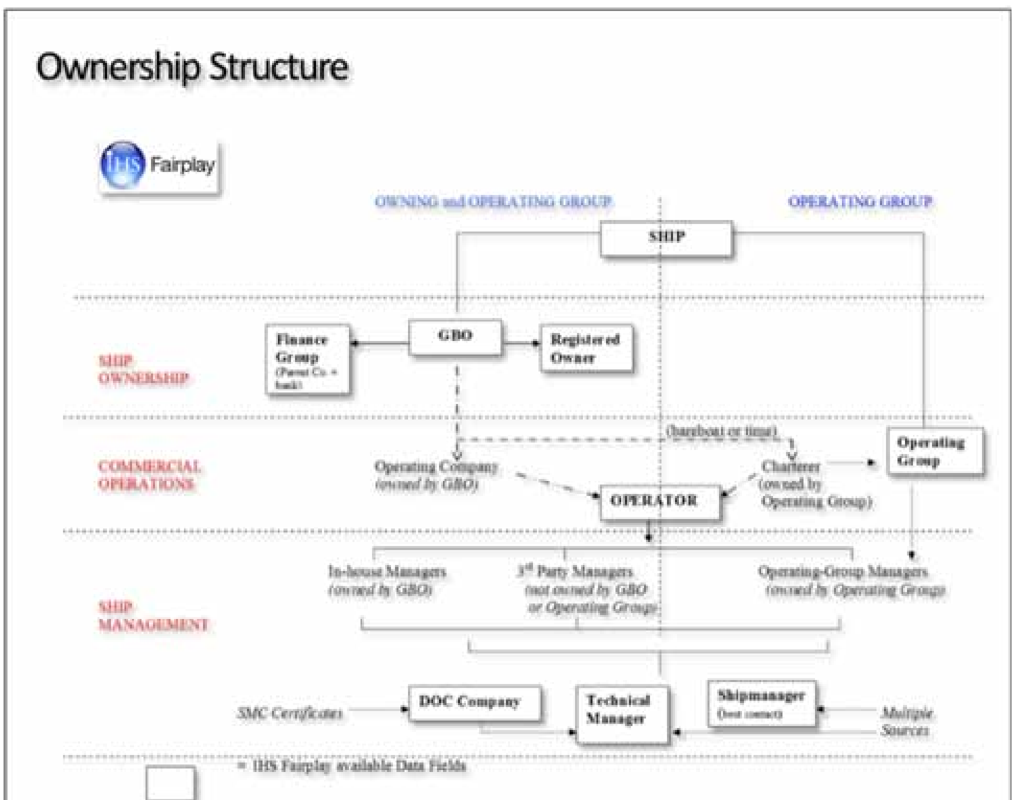
Mr. Michael Rodriguez, Deputy Administrator of the Maritime Administration (MARAD) of the U.S. Department of Transportation, facilitated Session VI and provided an introduction to the global supply chain topic. He stressed the importance of taking a whole-system approach and not solely focusing on maritime. In building communities of interest, collaboration, and considering broader uses of data, he specifically mentioned the Moving Ahead for Progress in the 21st Century Act (MAP-21) funding and authorization bill governing United States federal surface transportation spending and the provision to develop a national freight policy. Mr. Rodriguez also challenged the workshop participants to consider the distributed and decentralized nature of the maritime domain and the economic motives driving behavior.

Mr. Paul Kerstanski with the firm Pacific Architects and Engineers Incorporated (PAE) presented "Awash in a Sea of Data." As of May 2015, Mr. Kerstanski actively tracks more than 150,000 vessels every day. He provided insight into the world wide data stating three zettabytes (ZB) of data exist in the global universe today with expectations to grow to 40 ZB by 2020. Mr. Kerstanksi characterized a ZB in terms of 36 million years of HD video, and cited estimates that of the 3 ZB of data globally available today, 1 ZB is financially related. Considering the 240 billion annual financial transactions occurring globally, the ability of financial institutions' to manage this data is accomplished through automatically tracking, analyzing, and continuously monitoring shared information. The system is all about risk management and trust maintenance. Data mining of information such as credit history, bank accounts, property records, and spending habits feed risk assessments using various analytics tools and methods to monitor and validate each transaction. Looking at the maritime domain from this perspective, if information is shared, 500 million annual global maritime transactions globally could be continuously monitored with similar analytic techniques based on information such as vessel characteristics and history, flag state, ownership and management records, crew records, insurer information, and port inspection reports.

Mr. Paul Kerstanski

Mr. Kerstanski provided an overview of the complexity of vessel ownership and management, country relationships, parent companies, and the fleet. Key findings in seafarer studies show crew members from 123 countries were found on foreign-flag vessels calling at U.S. ports but only 10 nationalities made up the vast majority (77.9%) of crews. Overall, Asian countries supplied 59.4% of total crew members on foreign-flag vessels. Eastern European nations were the second greatest source of crew members at 22.1% of the total, and Western European nations were an important source of command officers (master and chief engineer). Notably, there is little relationship between vessel flag and nationality of crew members employed on the vessel.

In considering risk analysis, Mr. Kerstanski presented several indicators which might designate a vessel to be of interest including: high interest port of origin, enroute high interest port, transshipment port, degree of crew transparency (all crew members from a single country of interest), cargo transparency, trust-



ed conveyance (ship and containers owned by the organization of interest), high interest country of origin, and trusted shipper (customs form stamped by the armed forces, not customs).


Mr. Wayne Hoyle

Mr. Wayne Hoyle of MacDonald, Dettwiler, and Associates (MDA) presented on the use of satellite radar for maritime data analytics. The challenge of detecting potential threats and queuing high value assets spans many areas of concern to include sovereignty protection (foreign naval vessel in, or approaching territorial waters), counter-piracy (possible pirate vessels and historic piracy events), counter-terrorism (potential threats from inbound vessels), fisheries protection (vessels engaged in illegal fishing within EEZs or controlled areas), environmental protection (vessels engaged in illegal bilge dumping, accidental oil spills), and counter-trafficking (suspicious inbound vessels, or vessels traversing territorial waters). The MDA BlueHawk system integrates data derived from the high-resolution RADARSAT satellites, commercial optical satellites, and aerial systems.

Mr. Hoyle presented three case studies illustrating the use of the MDA BlueHawk system: human trafficking, counter-piracy, and U.S.-based maritime domain awareness. Highlighting the magnitude of migrant rescues in the Mediterranean, Mr. Hoyle walked through a specific example of vessels detected off the coast of Algeria on September 21, 2014. With 2 of the 24 radar-detected vessels not correlated with AIS, the BlueHawk system identified 1 of the dark targets to be a 141-m cargo vessel likely to be legally trading. The second dark target, a 24-m vessel, was transiting in a known smuggling route and triggered a system report to notify local authorities to queue a high value asset. Mr. Hoyle referenced the International Maritime Bureau report of April 22, 2015, stating that Global Piracy is up 10% from the previous year with over 55% of attacks in South East Asia and no incidents in Somalia in Q1 of 2015. Mr. Hoyle specifically walked through the incident of the Vietnamese tanker, MT SUNSRISE 689 that had gone missing off the coast of Singapore in October 2014. There were 31 RADARSAT-2 scenes collected from October 1st to 15th detecting over 1,500 vessels in the region. Using MDA BlueHawk, by filtering on vessel size and time, the field of 1,500 vessels was reduced to 95 possible vessels. Using the MT SUNSRISE 689's last known position and the maximum speed of the tanker to determine the possible range, analysis of the radar image reduced the identification to the single vessel of interest. The last case study on U.S. based MDA illustrated the RADARSAT-2 collection with coverage of a stretch of 1,300 nautical miles imaged along the Pacific coast (175,000 square nautical miles) and 1,150 nautical miles along the Atlantic coast (920,000 square nautical miles). A single scene off the Pacific coast yielded 80% to 90% correlation of radar and AIS. Within that scene, Mr. Hoyle walked through the methodology to determine which dark targets are vessels of interest.

Lastly, Mr. Hoyle walked through the next steps for MacDonald, Dettwiler, and Associates to include: support for additional synthetic aperture radar (SAR)/optical satellites, improved radar and AIS correlation, increased analytics based on AIS, vessel registry, SAR, and other data sources, moving from "SAR detection" to "SAR classification," and reduce latency from SAR image collection to vessel reporting to less than 30 minutes. Additionally, in response to the 2014 GMF, Mr. Hoyle promoted the release of sample data for experimentation and development.


Mr. Jim Pietrocini

Mr. Jim Pietrocini of the company SAP ("Systems, Applications, and Products in Data Processing") presented the topic "Maritime Machina." SAP is the 4th largest software company, behind Microsoft, Oracle, and IBM, with systems touching 74% of the world's transaction revenue. Known for its Enterprise Resource Planning (ERP) business management software, it now operates in five market categories including software applications, analytics, mobile applications, cloud technology, and database technology. Its specific solution for big data is the SAP HANA computing and analytics platform.

Mr. Pietrocini discussed several areas of innovation and change including the use of crowdsourcing, machine learning, and transformational natural language processing. Mr. Pietrocini depicted the "No Man's Sea" with the seas shrinking. As missiles grow longer-ranged and more precise, as sensors grow ever sharper, there are ever fewer places for a ship to hide. He also shared a future "at sea" perspective including: the proliferation of anti-ship cruise missiles (ASCMs), increased lethality of smaller vessels and shore batteries, relatively poor staying power of larger expensive vessels, Ka-band connectivity and 4G, and data scientists as new operations specialists. Lastly, Mr. Pietrocini described the impact of Open Source Intelligence (OSINT) and Social Intelligence (SOCINT) operations and with Global Command and Control System - Maritime (GCCS-M) over 20 years old, the need to rebuild Command, Control, Communications, Computers, and Intelligence (C4I) systems.

In a time of profound change for U.S. national security, Mr. Pietrocini outlined the following challenges requiring a closer Department of Defense (DoD) and industry relationship and revolutionary change: a requirement to innovate and advance the mission while trimming costs, floods of structured and unstructured data requiring efficient analysis and management, mission teams fighting a constant barrage of evolving cyber treats in real time, a growing need to fuse and share data across platforms and security levels, and the instant access and insights demanded from the analyst by the operator.

Discussion

The 2015 Global Maritime Forum was sponsored by RADM Train (Director, NMIO) and hosted by NASA Ames Research Center in Mountain View, California. Attended by approximately 120 people from government, academia, and the commercial sector, this workshop focused on data analytics for making sense of maritime activity. The theme (Turning the Corner in the Maritime Domain: Leveraging Data to Achieve Effective Understanding) was well described by the opening remarks of RADM Train and keynote speakers.

RADM Train emphasized the importance of embracing new analytic methods and new collaboration partners to address maritime security on a whole-of-nation basis, including government, commercial, and academic sectors. Her remarks focused on the importance of improved understanding to inform decision-making and clearly pointed to the national security implications of trends that are not traditionally considered military, citing the example of capacity-challenging levels of illegal migration in the Mediterranean and concerns for this being an avenue for terrorists to exploit for gaining entry into Europe.

Dr. Eng Lim Goh, Chief Technology Officer at Silicon Graphics Inc., presented a stunning array of high performance computing renderings (drawn from computational fluid dynamics) and data intensive computing ("big data analytics") examples. He stressed the imperative to avoid eliminating or thinning data and the importance of good visualizations to help discover what information the data contains. His talk also addressed machine learning (Deep Blue, Watson, Pokerbot) and the critical importance of noticing disparate data or what might otherwise be considered noise. Mr. Kshemendra Paul, PM-ISE, spoke about distributed, decentralized, and coordinated information sharing.

Key enablers include format standardization (for example, the National Information Exchange Model, NIEM) and a recognized level of information assurance and access control.

The presentations spanned a range of technical disciplines including Platforms and Simulations, Data Mining and Analytics, Data Visualization and Accessibility, IUU Fishing, Interdiction and Safety, and the Global Supply Chain. Each session has specific implications for the maritime domain. As part of the collaboration sessions at the workshop, participants were challenged to design a data competition drawing on maritime data (including vessel positions, economic data, business interests, and other sources) to address any one of several maritime issues such as IUU Fishing, illicit trafficking, and national security. These data competition recommendations are valuable input into the planning of a maritime domain awareness data challenge.

Session Findings

The **Platforms and Simulations** session made clear that the ability to collect, store, and analyze massive amounts of data is at hand. Persistent earth observation from space is enabling completely new disciplines of analysis, currently focused (in the commercial world) on global-scale financial, economic, and environmental issues. Near-term trends will include global persistent earth observation at commodity prices, with resolution adequate for many important analyses. New technologies in space (for example, software defined radios and multispectral imagers) will bring products to market that were once available only to governments (i.e., ELINT and IR imagery). The role of these new sources of data and their availability to friend and foe alike should be considered on at least three planes: improving operational and cost effectiveness; the role of commercial space in our national architecture; and the impact of commercially available information on our security risk environment (with particular attention to Anti-access, Area Denial strategies).

The **Data Mining and Analytics** session made clear the importance of operating on a wide range of data types to derive understanding and competitive advantage. Interdisciplinary collaboration can bring these data sets and the computational capabilities to bear on very difficult problems. Our doctrinal view of sense-making from vessels, cargo, people, and infrastructure is on the right track, but lacks many other important drivers such as social, economic, and environmental forces. It is time to encourage novel analytic approaches, to find ways to operate on data sets that are currently too closely held to be available and to more actively engage the commercial and academic sectors in assuring maritime security. New methods, drawn from artificial intelligence, machine learning, and social network disciplines may be enablers for complex problems of risk assessment.

The **Data Visualization and Accessibility** session highlighted interagency and international collaboration, enabled by well-defined standards ("making data accessible"), low barriers to entry and intuitive user interfaces ("making it easy"), and presenting analytic results ("visualization") to trigger human insights and intuition. It is important that we reach outside of U.S. Government channels for creative insights and intuition regarding new analytic processes.

The **IUU Fishing** session recognized the fundamental importance of controlling fisheries to ensure long-term world nutrition and addressed the difficult problems of detecting, attributing, deterring, and reducing IUU fishing. Speakers highlighted technical successes and shortcomings both at the analytic level and at the operational level. Massive, multi-dimensional data sets proved valuable in predictive analytics, and multi-layered operations useful in enforcement. The enormity of the problem makes clear that even limited success will require considerable effort at all levels of society, from scientific research through operations to national and regional policy-making. It is also clear that IUU fishing undermines efforts to ensure sustainable living marine resource levels and triggers regional instability. This, in turn affects our national security posture. It is time to develop competencies in the MDA community that broaden analytic capabilities beyond vessel tracking and into analytic products, derived periodically for every known vessel from global and pervasive data sets that better inform command decision-making.

The **Interdiction and Safety** session explored the role of direct information and intelligence sources, automated analytic processes, and hybrid approaches for cuing maritime threat operational responses. Experience shows that random patrolling is ineffective when compared to cued enforcement operations. It also shows that abnormal behavior is not always associated with illicit activity, but that including analysis of the entire supply chain may be required. This session suggests that illicit activity is a normal part of maritime activities, and that isolating the maritime segment of a commercial transaction may miss most or all of the important cues. Attacking the problem by in-depth computer-assisted analysis of individual sectors of maritime activity may make the problem tractable.

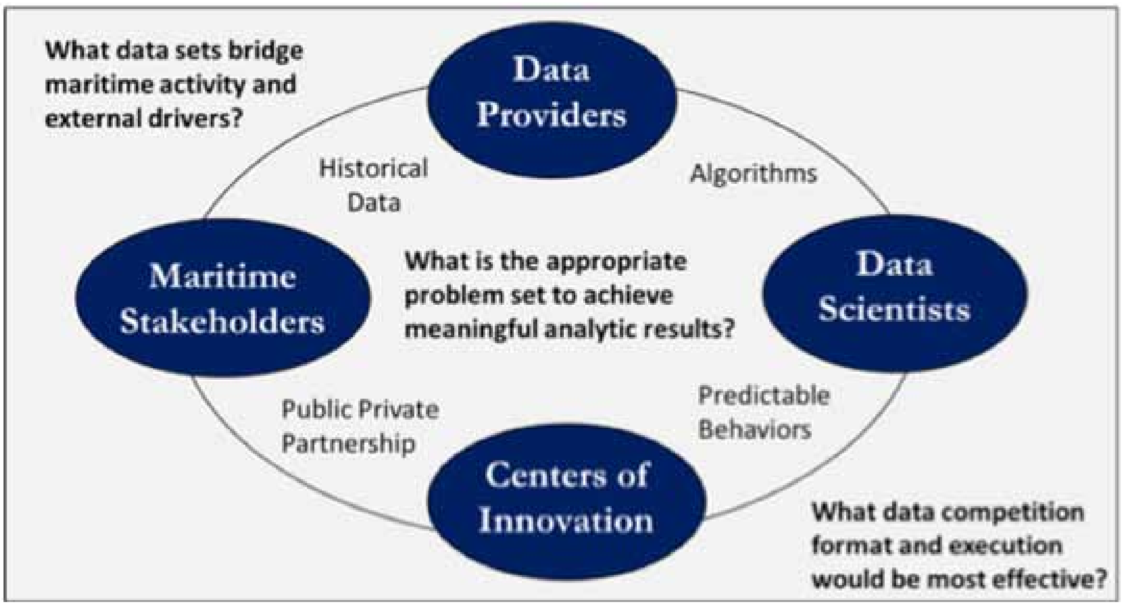
The **Global Supply Chain** session considered the role of analytics in enabling legitimate use of the maritime domain. Risk analysis and threat evaluation are critical components, which build on vessel positions and tracking by adding commercial, environmental, social, and other factors. The current state of the art is in business rule-based scoring systems; there is room for machine intelligence to fuse more varieties of input variables and to create more complex evaluation algorithms. Local knowledge is essential to understanding the context of maritime activity; encoding this to make it globally available is possible with business rule-based systems. Mobile computing and information systems, as well as natural language processing capabilities, are poised to play a new and important role in supply chain analysis. The challenge for analytic processes is achieving "effective understanding" to better inform command decision-making. This is impossible without considering parts of the supply chain outside of the maritime sector and impossible without including input variables that cross current interagency title authority lines. There are policy challenges and technology challenges; change will only be motivated by demonstrating value-added by a new generation of analytics, and for practical reasons, teaming with the commercial and academic sectors should be a priority.

Acknowledgement: Session findings were contributed by Dr. John Mittleman of the Naval Research Laboratory.

Collaborations Sessions

This year's sessions strove to foster the collective wisdom and expertise of GMF participants and sketch out a data competition for maritime domain awareness. This idea was born from the 2014 GMF recommendations focused on data analytics and the need for algorithms, automating analysis, and determining predictable behaviors.

Work groups were constructed with a broad representation of the maritime domain and external factors, data providers/owners, scientists, and centers for innovation. To facilitate the collaborative sessions, the following documents were provided to participants: scoping of MDA, examples of three data competitions (NASA, GE Flight Quest I and II, and Heritage Health Prize), and worksheets to guide discuss and collect input.



Schedule

<u>Day 1 Collaboration Session (1 hour) – Familiarization of the Domain</u>

- Brief introduction to the session. Task: Focus on first question:
- "What data sets bridge maritime activity and external drivers?"
- Goal: Familiarization of MDA across the team and identifying dependencies or external drivers.
  Begin discussion on "What is the appropriate problem set to achieve meaningful analytic results?"

<u>Day 2 Collaboration Session (1 hour) – Designing a Data Competition</u>

- Brief introduction to the session. Task: Focus on questions:
- "What data competition format and execution would be most effective?
- Goal: Determine a problem to be solved and the associated data competition format.

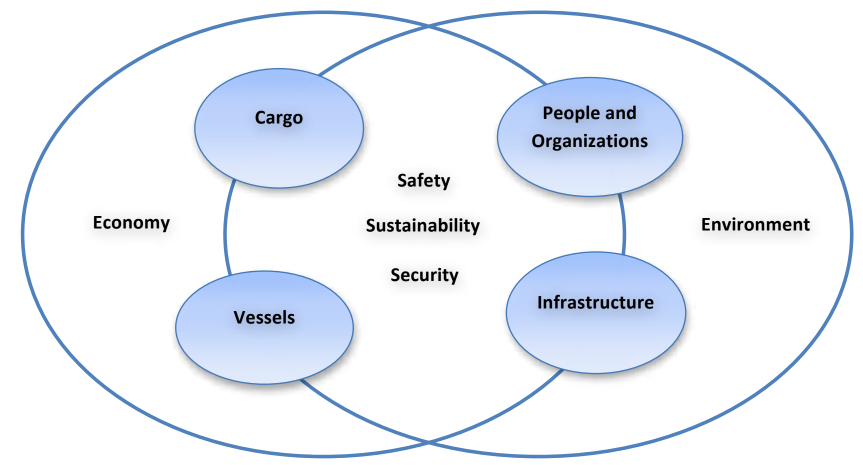
<u>Collaboration Session Synthesis (1 hour) – Brief out by each team</u>

Scope of Maritime Domain Awareness (MDA)

The scope of MDA is defined as follows:

- Maritime Domain Awareness (MDA) is the effective understanding of anything associated with the maritime domain that could impact the security, safety, economy, or environment of the United States.

- The Maritime Domain is all areas and things of, on, under, relating to, adjacent to, or bordering on a sea, ocean, or other navigable waterway, including all maritime-related activities, infrastructure, people, cargo, vessels, and other conveyances.



Complexity of the Maritime Domain and Independency with the Economy and the Environment

MDA information involves a broad range of data categories which may include:
• Vessels—information such as flag, type, classification society, tonnage, maximum speed, origin, positional information, next port of call, last port of call, track history, construction and outfitting, history (build, employment, and regulatory), documentation, acoustics, capacities, etc.

• Cargo—information derived from cargo manifests and bills of lading, including characteristics, origin, handling instructions, destination, and hazard class, as well as information derived from customs and hazardous material inspections; chemical, biological, nuclear, radiation, or explosive detection sensors; and data exchange and mandatory reporting systems.

• People and Organizations—information regarding vessel owners and charterers, crew and passengers, freight forwarders, husbanding agents, insurers, lien holders, port terminal operators, stevedores, etc., as well as financial transactions that people and organizations may be involved in that indicate whether relationships are legitimate, illicit, or demonstrative overt or covert activity.

• Infrastructure—information with maritime attributes, including requisite geospatial information, such as the following:
  - o Ports, Waterways, and Facilities - piers, terminals, cranes, fueling facilities, and other resources or key limits, such as vessel traffic services and vessel separation schemes, shipping and great circle routes, international maritime boundaries, disposal sites, offshore leasing sites (e.g., oil fields, wind farms, and other national energy security components), etc.
  - o Critical Infrastructure - locks, bridges, tunnels, channels, aids to navigation, undersea cables, pipelines, nuclear and other power plants, and intermodal connections.

External Factors/Interdependencies:
• Environment—information, data, and metadata on weather including wind, sea, swell, tides and currents, other hydrographic and bathymetric data, sea temperature and salinity, and ice flows, as well as information regarding maritime natural resources, regulated fisheries, migratory patterns, marine sanctuaries, marine protected areas and species, pollution, emission control areas, and impacts from offshore energy development, etc.

• Economy—information, data, and metadata on the production, distribution or trade, and consumption of limited goods and services. Such data includes Gross National Product and its components, Gross National Expenditure, Gross National Income in the National Income and Product Accounts, and also the capital stock and national wealth. Other economic indicators include a variety of alternative measures of output, orders, trade, the labor force, confidence, prices, and financial series (e.g., money and interest rates). At the international level, there are many series including international trade, international financial flows, direct investment flows (between countries) and exchange rates, etc.

Worksheet

- **Title:** (Title of the competition and problem statement)

- **Proposed Activities:**
  High-level summary of the competition (what problem is being solved, what are the competition stages/phases, participants, cost, timeline, etc.).

- **Proposed Justification:**
  Address why it is important/what is the short and long-term impact.

- **Success:**
  What would a winning entry actually look like (form)? What criteria would be used to judge such an entry (what should success: look like?)

- **Metrics and Judges:**
  What metric(s) should be used to evaluate the winner(s)? Are the success criteria sufficiently known and quantitative to allow for an automated approach to judging winning entries or is human judgment required? What should the process be for judging the winning entry(ies)? Would verification tools or demonstration events be required for unbiased comparative performance assessment of the solutions?

- **Eligibility and Participants:**
  Who would be the target population (entrants) for the proposed competition? Would there be restrictions on who could compete and what would they be?

- **Incentives:**
  What incentives would be optimal to attract the desired entrants and produce the desired results of the competition? Should there be multiple prizes? Are there any non-monetary awards or benefits that should be considered? How large does the prize purse need to be to stimulate investments in desired technologies?

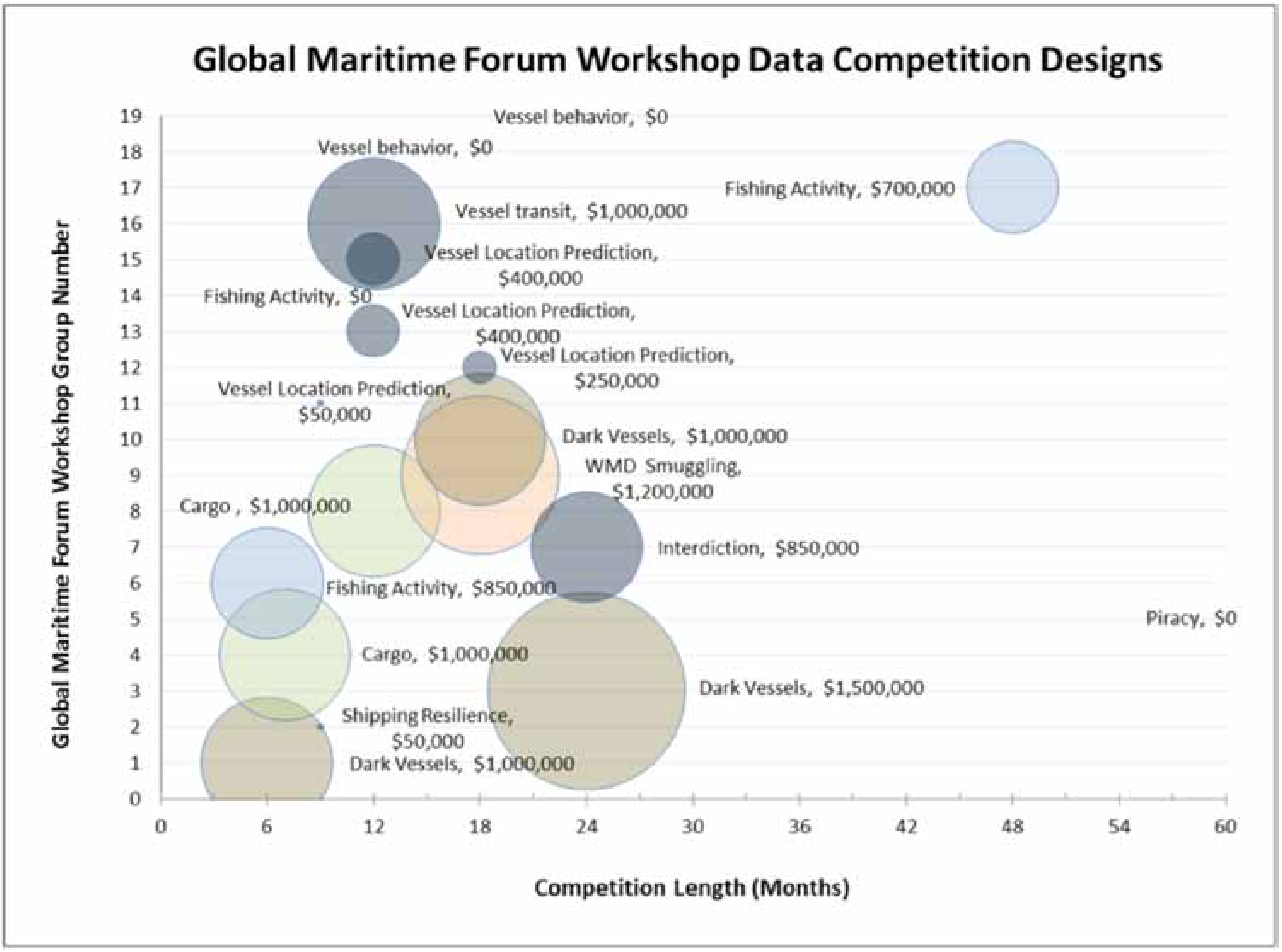
- **Timeline:**
  How long should the competition run? Should the competition be run in stages? If so, how long for each stage? What is an estimated timetable to prepare and launch the competition?

- **Partners:**
  List partners below that might have an interest in addressing the problem statement. What would those partners bring to the table?

- **Communications and Outreach:**
  How would the competition structure communicate with contest entrants and potential participants (website? social media? leaderboard? interactive forums? etc.)

Maritime Domain Data Challenge Recommendations

The GMF workshop leveraged the robust and diverse participation of more than 120 U.S. and international stakeholders and subject matter experts from academia, government, commercial maritime, and Silicon Valley. Dedicated collaboration sessions with 19 pre-assigned teams were established to identify a problem in the maritime domain that could be solved using advanced data science in the format of a data competition. The resulting 19 recommendations for a maritime domain data challenge varied greatly in terms of scope and the particular mission focus; from $1M prizes to automate the identification of dark vessels to identifying and predicting vessel behavior without monetary award. The general themes, duration of the proposed challenge, and potential prize amount are visually displayed in the following chart:



Group 1 identified the problem of minimizing unidentified "dark targets" or non-emitting vessels using open source or commercial data. The goal of this competition is to establish a standard approach (taxonomy) to identify dark targets and achieving an identification success rate with 80% accuracy.

Group 2 focused on shipping resilience with a natural disaster scenario in San Francisco. The goal of the competition would be to minimize economic loss during the shutdown of a major port.

Group 3's challenge "Shining the Light on Dark Targets" focused on how to detect dark targets, i.e., boats that are not emitting signals. The contest would be a phased approach of soliciting proposals, data analysis, and tool development. The goal of this competition is to detect dark targets with 95% confidence within an hour time frame to be tactically relevant.

Group 4 focused on the identification of illicit cargo with the goal of improving anomaly detection for illicit cargo identification and prioritization for search and seizure. Two years of data is proposed for the contest including: AIS, search and seizure, container tracking, and bill of lading. Success would be measured by the improved accuracy of identification of illicit cargo.

Group 5 identified the challenge of developing a predictive system to combat piracy events and predict future piracy events. Using historical data and current data, the goal of the contest is to increase the accuracy of prediction in comparison to actual piracy events.

Group 6 focused on the ability to predict the behavior of fishing vessel traffic. Using a staged approach, the goal of prediction accuracy would be aligned to the time frame of 50% accuracy at 3 months, 80% at 1 month, and 90% at 1 week.

Group 7 identified a challenge of identifying vessels with operational concerns in safety, sustainability, and security to improve vessel interdiction. Using a staged approach of prototyping and tool development, the goal is to increase the interdiction rate of success at key ports.

Group 8 identified the need to improve cargo container accountability. The goal of this challenge would be to increase cargo container manifest data veracity for port and state use. TEU (twenty-foot equivalent) contents would be tested on demand by port staff, at designated waypoints within supply chain, demonstrating 50% reduction in inconsistencies.

Group 9 identified a Weapons of Mass Destruction (WMD) smuggling exercise with the goal of determining the greatest vulnerability to a WMD smuggling attack.

Group 10 focused on dark target identification using multiple Unmanned Aerial Systems (UAS) platforms and sensors. Participants would develop algorithms to analyze multiple UAS sensor data, correlate with other vessel data sources to either ID, detect, and classify. Measured by timeliness and accuracy, using UAS data, the goal is to find the best way to provide operators with vessels of interest not self-reporting.

Group 11 identified a vessel location prediction challenge with the goal of identifying the highest number of vessels within a defined area (10 km by 10 km). The challenge would be measured by the success of prediction at 3 time frames: 4 hours, 12 hours, and 24 hours.

Group 12 identified a challenge to predict the location of industrial shipping vessels (fish carriers) based on historical data. Using a phased approach, participants would identify and assemble data sources, develop a predictive model and indicators for suspicious activity, and develop pressure points for influencing behavior. The success of the top predictive models would be judged based on the behavior of fish carriers in 2016.

Group 13 focused on the prediction of maritime incidents with a goal of enabling law enforcement and border protection service to resource and position assets in a time and place to detect and respond. The challenge would be measured by the accuracy of event predictions based on observed and available data.

Group 14's maritime patron optimization challenge aims to use data science to increase "on target" Marine Protected Area (MPA) flights to gather evidence of IUU fishing. The participants would identify patterns of fishing activity within certain zones by developing algorithms against known positional data. The goal is to have the highest percentage of true positive identifications to false positives.

Group 15 identified a vessel location prediction challenge with the goal of achieving 95% accuracy. The contest would be performed in multiple stages.

Group 16 focused on predicting deviations from vessel transit routes to improve the efficiency of the marine land shipping system. Participants would incorporate robust data streams to produce an accurate model. The predictive model would be accessed for accuracy over a 6-month period of activity.

Group 17's challenge, Scaling Back, focuses on the sustainability of fish stocks and measuring fishery activity. Participants would predict the value of key fisheries metrics for specified countries to include fishing quotas, prices, quantity landed, and size of registered fishing fleet. Using historical data as ground truth, the data competition would be executed in phases with predictions for 2018, 2019, and 2020. The prediction would be measured against the actual values with a goal of achieving the lowest error.

Group 18 formulated the challenge, Seafinder, to predict vessel patterns for any location. Based on historical data, the goal of the challenge is to use diverse data sets to predict vessel activities for various locations to inform enforcement resources. The predictions would be evaluated to achieve 85% accuracy.

Group 19 identified a challenge to predict illegal behavior at sea based on activity on land with a goal of improving enforcement efficiency. By assembling low cost data of activities, income, affiliation, and linking the data to vessels, the participants would predict the likelihood of illegal activity and rank vessels by infraction probability. From a sample of enforcement activities (e.g., 100 boardings), the challenge would use actual citation rates to evaluate the success rate of predicting vessels cited for infractions.

Agenda

While there have been incremental steps taken by governments and the maritime industry to move forward in collecting, fusing, and visualizing data relevant to operations and security in the maritime domain, achieving true awareness has been elusive. At the same time, currently available technology being used across a broad spectrum of industries for analyzing and understanding complex, "unpredictable," and big data sets is progressing by leaps and bounds. Turning the corner in the maritime domain will require innovative thinking and unique collaboration to advance to the next art of the possible.

Day One presentations will highlight available and emerging technologies analyzing and understanding complex data sets and environments. Day Two will focus on effective understanding, sharing information to achieve a holistic view of this complex environment, and understanding how advanced analytics can be applied to specific behaviors within the maritime domain.

Pre-Workshop Social

Monday, 1 June	
Time	**Event**
1900–2100	**Social Event at the Computer History Museum** Sponsored by MDA Corporation, Planet OS, SAP® (Early Registration)

Day 1

Tuesday, 2 June	
Time	**Event**
0730–0830	**Registration & Morning Coffee**
0830–0845	**Welcome Remarks** Dr. Paul Shapiro, *Science Advisor, National Maritime Intelligence-Integration Office*
0845–0915	**Welcome Address** Dr. Eugene Tu, *Center Director, NASA Ames Research Center* Rear Admiral Elizabeth Train*Director, National Maritime Intelligence-Integration Office; Commander, Office of Naval Intelligence*
0915–0945	**Keynote** Dr. Eng Lim Goh, SGI
0945–1000	**Break**
1000–1115	**Session I: Platforms and Capabilities.** The latest developments and achievements in high-performance computing (HPC), grid computing, and Cloud and High Scalability Computing (HSC). Chair: Dr. Piyush Mehrotra, *Ames Research Center (ARC)* Speaker 1: Mr. David Pellerin, *Amazon* Speaker 2: Mr. Rainer Sternfeld/Dr. Christopher Clark, *Planet OS* Speaker 3: Mr. Justin Langlois, *Google/Skybox*
1115–1130	**Break**
1130–1245	**Session II: Data Mining and Data Analytics.** Emerging techniques and breakthroughs in the extraction of patterns and knowledge from vast amounts of data and methods for large-scale data analytics. Chair: Dr. Paul Shapiro, *NMIO* Speaker 4: Mr. Alan J. Broder, *Novetta Solutions* Speaker 5: Dr. Sangram Ganguly, *NASA Ames Research Center* Speaker 6: Dr. Rodney Martin, *NASA Ames Research Center*
1245–1300	**Workshop Photo**
1300–1400	**Lunch**
1415–1545	**Session III: Data Visualization and Accessibility.** Realizing the potential of Big Data by advancing the art and science of making complex data more accessible, understandable, and usable. Chair: Dr. Sean Everton, *Naval Postgraduate School* Speaker 7: Mr. John Stastny, *SPAWAR* Speaker 8: Dr. Chris Henze, *NASA Ames Research Center* Speaker 9: Dr. Alonso Vera, *NASA Ames Research Center*

Tuesday, 2 June (cont.)	
1545–1600	**Break**
1600–1700	**Parallel Collaboration Session, Day One**
1700–1730	**Wrap up and Adjourn**

Day 2

Wednesday, 3 June	
Time	**Event**
0800–0830	**Morning Coffee**
0830–0845	**Administrative Remarks**
0845–0900	**Day 2 Introductory Speaker**
0900–0915	**Day 2 Keynote** Mr. Kshemendra Paul *Program Manager, Information Sharing Environment*
0915–1030	**Session IV: IUU Fishing and the Maritime Environment.** Current challenges from a policy, operations, and technology perspective and the future implications for maritime security. Chair: Dr. John Mittleman, *U.S. Naval Research Laboratory* Speaker 10: Ms. Jenifer Austin, *Google* Speaker 11: Mr. Todd Jacobs, *National Oceanic and Atmospheric Administration (NOAA)* Speaker 12: Dr. Robert Nolker, *Analyze Corp.*
1030–1045	**Break**
1045–1200	**Session V: Interdiction and Safety.** Current and future challenges of maintaining the safest possible maritime environment and improving the interdiction of illicit flows of goods and people. Chair: Mr. Scott Genovese, *Global Maritime Operational Threat Response Coordination Center* Speaker 13: Ms. Karna Bryan, *NATO Centre for Maritime Research and Experimentation (CMRE)* Speaker 14: Mr. Neil Watts, *United Nations* Speaker 15: Dr. Harm Greidanus, *European Commission – Joint Research Centre (JRC)*
1200–1300	**Lunch**
1300–1415	**Session VI: Global Supply Chain.** Understanding the movement of cargo, vessels, and people while preserving the efficiency and effectiveness of the shipping industry. Chair: Mr. Michael Rodriguez, *U. S. Department of Transportation* Speaker 16: Mr. Paul Kerstanski, *A-T Solutions* Speaker 17: Mr. Wayne Hoyle, *MDA Geospatial Services* Speaker 18: Mr. James Pietrocini, *SAP*
1415–1430	**Break**
1430–1530	**Parallel Collaboration Session, Day Two**
1530–1545	**Break**
1545–1645	**Brief Outs**
1645–1700	**Closing Remarks**

Forum Participants

Rear Admiral Elizabeth Train, US Navy
National Maritime Intelligence-Integration Office
Office of Naval Intelligence
Co-Host

Dr. Eugene L. Tu
NASA Ames Research Center
Co-host

Mr. Dean Ahner
NCTC / ODSI / Data Science and Innovation Group
Participant

Ms. Midori Akamine
World Bank
Participant

Professor Craig Allen
University of Washington
Participant

Mr. Nick Andersen
US Coast Guard Intelligence CIO
Participant

Ms. Jenifer Austin
Google
Speaker, Facilitator

Ms. Edith Backman
National Maritime Intelligence-Integration Office
Participant

Mr. Jatin Bains
Channel Logistics
Participant

Mr. Brent Baker
Channel Logistics
Participant

Mr. Nick Baranishyn
Strategic Analysis Branch Los Angeles Joint Regional Intelligence Center
Participant

Mr. Tony Bauna
KONGSBERG Satellite Services AS (KSAT)
Participant

Ms. Tamaki Bieri
Stanford University
Participant

CDR Kurt Birkhahn, US Navy
DoT, Maritime Administration
Participant

Mr. Kevin Blackwell
Metris
Participant

Mr. Bill Bonwit
SPAWAR
Participant

Mr. Terry Boone
NOAA Fisheries Office of Law Enforcement, Pacific Islands Division
Participant

Dr. Alex Bordetsky
Naval Postgraduate School
Facilitator

Dr. Mark Brady
National Marine Fisheries Service (NMFS)
Participant

Mr. John (Curt) Bransfield
Joint Interagency Task Force West (JIATF West)
Participant

Mr. Alan J. Broder
Novetta Solutions
Speaker

Mr. Malcolm Brown
UK National Maritime Information Centre
Participant

LT Candace Brueggeman, US Navy
Seal Delivery Vehicle Team One
Participant

Ms. Karna Bryan
NATO Centre for Maritime Research and Experimentation (CMRE)
Speaker, Facilitator

Mr. Rob Burkhard
Riptide Technology, Inc
Participant

CAPT David Burns
Royal Navy N2, Maritime Operations Centre, Northwood HQ
Participant

CTRC Vincent Campolo, US Navy
Naval Special Warfare Mission Support Center
Participant

Mr. Alexander P. Carrillo
Orange County Intelligence Assessment Center
Participant

Dr. Chung-Fu Chang
Johns Hopkins Applied Physics Laboratory
Participant

Ms. Lynn Collins
SPAWAR
Participant

LTCOL Dean Commons
Australian Customs and Border Protection Service
Participant

Mr. Antoine de Chassy
Spire
Participant

Dr. Adarsh Deepak
Science and Technology Corporation (STC)
Participant

Mr. Daren Devlin
USCG Intelligence Coordination Center
Participant

Mr. Peter Dorcas
exactEarth
Participant

Mr. Andy Estep
Waitt Institute
Participant

Dr. Sean Everton
Naval Postgraduate School
Session Chair, Facilitator

Mr. Robert Farrell
CA Fish and Wildlife
Participant

LT Brian Ferguson, US Navy
Naval Special Warfare Group THREE
Participant

Mr. Rory Fitzpatrick
National Space Centre
Participant

Mr. David Floyd
Draper Laboratory
Participant

CDR Joseph Fraser, US Navy
Commander Naval Forces Europe & Africa / Commander 6th Fleet
Participant

Mr. Letterio Gemelli
Thales Alenia Space
Participant

Mr. Scott Genovese
Global Maritime Operational Threat Response Center Coordination Center
Session Chair, Facilitator

LTJG Adam Gibbs
US Coast Guard
Participant

Dr. Uwe Glässer
Simon Fraser University
Participant

Dr. Eng Lim Goh
SGI
Keynote Speaker

CDR Michael Goldschmidt
US Coast Guard
Facilitator

Dr. Harm Greidanus
European Commission - Joint Research Centre (JRC)
Speaker, Facilitator

Ms. Carol Hayes
National Space Centre
Participant

Dr. Eric Hendricks
SPAWAR
Participant

Dr. Chris Henze
NASA Ames Research Center
Speaker

Ms. Erin Hill
National Geospatial-Intelligence Agency
Participant

Mr. Ken Holliday
Office of the Program Manager - Information Sharing Environment
Participant

Mr. Greg Hosking
Australian Customs and Border Protection Service
Participant

Mr. Wayne Hoyle
MDA
Participant, Sponsor

Mr. Todd Jacobs
NOAA Unmanned Aircraft Program
Speaker

LCDR Maureen Johnson
US Coast Guard
Facilitator

Mr. Paul Kerstanski
A-T Solutions
Speaker

Mr. David King
US Coast Guard Pacific Area Command
Participant

Mr. Koichi Kishi
NEC
Participant

Mr. Hideshi Kozawa
NEC
Participant

Mr. Justin Langlois
Google - Skybox
Speaker

Mr. Edward Lau
MDA Coorporation
Participant, Sponsor

Mr. A. Lyston Lea II
National Maritime Intelligence-Integration Office
Organizer

Dr. Kai N. Lee
David & Lucile Packard Foundation
Participant

Ms. Hi Dee Lehnert
Air and Marine Operations Center (AMOC)
Participant

Ms. Cassandra Lopez
Office of Naval Intelligence
Facilitator

Mr. Andrew Loretta
ORBCOMM
Participant

Mr. Mark E. Louttit
Analytical Support Massachusetts Commonwealth Fusion Center
Participant

Mr. Trent Lukaczyk
Stanford University
Participant

Lt(N) Andrew MacKay
NORAD
Participant

Dr. Rodney Martin
NASA Ames Research Center
Speaker

Dr. Phil McGillivary
US Coast Guard
Participant

LCDR Carrie McKinney
US Coast Guard
Facilitator

Dr. Piyush Mehrotra
NASA Ames Research Center
Organizer, Session Chair

Mr. Gregory Melcher
Johns Hopkins Applied Physics Laboratory
Participant

Mr. Joseph Milligan
MDA Deputies/NMCO/MSIPC
Participant

Mr. Benjamin Minichino
CLS America
Participant

Dr. John Mittleman
Naval Research Laboratory
Organizer, Session Chair, Facilitator

Ms. Gina Morello
NASA Ames Research Center
Organizer

Dr. Larry Morgan
Pacific Northwest National Laboratory
Participant

Mr. David Mottarella
Harris Corporation
Participant

Mr. Peter Michael Nielsen
Danish Embassy
Participant

ME5 Choon Hock NIOW
Republic of Singapore Navy
Participant

Mr. Steve Nixon
Vulcan Aerospace
Participant

Dr. Robert Nolker
Analyze Corp.
Speaker

Mr. Peter Nurney
Liquid Robotics Inc.
Participant

SLTC Nathan Ong
Republic of Singapore Navy
Participant

Mr Michael Pagnotti
Joint Interagency Task Force West (JIATF West)
Participant

Lt(N) John Paquette
NORAD-USNC NJ3
Participant

Mr. Kshemendra Paul
Office of the Program Manager - Information Sharing Environment
Keynote Speaker

Mr. George Peacock
Johns Hopkins University Applied Physics Laboratory
Participant

Mr. David Pearl
NOAA
Facilitator

Mr. James Pietrocini
SAP
Speaker, Sponsor

Mr. Steven Plystak
Harris Corporation
Participant

RADM (Ret.) Richard Porterfield, USN
Institute for Defense Analyses (IDA)
Participant

Professor Emeritus Vaughan Pratt
Stanford University
Participant

CAPTAIN(N) William Quinn
NORAD-USNORTHCOM
Participant

LT Jedediah Raskie
US Coast Guard Pacific Area Command
Participant

Mr. Guy Ravine
Revolutionary Labs
Participant

Mr. Joshua Reiter
Office of Naval Intelligence
Facilitator

Mr. Mark Richardson
The Pew Charitable Trusts
Participant

Mr. Randy Ridley
Planet OS
Participant

Dr. Richard Rikoski
Hadal
Participant

Mr. Michael Rodriguez
US Department of Transportation
Speaker

Mr. Ted Schmitt
Vulcan Inc
Participant

CMDR A.C. Schroder
RAN
Participant

Dr. Landon H. Sego
Pacific Northwest National Laboratory
Participant

Dr. Paul Shapiro
National Maritime Intelligence-Integration Office
Organizer, Session Chair

Ms. Kate Sherman
Anthropocene Institute
Participant

Ms. Catherine Simons
US Navy
Participant

Mr. Chandler Smith
COM DEV USA (exactEarth USA)
Participant

Professor João Sousa
Porto University
Participant

Mr. John Stastny
Space and Naval Warfare Systems Center Pacific
Speaker, Facilitator

Mr. Rainer Sternfeld
Planet OS
Speaker, Sponsor

Mr. Mark Stoddard
Defense Research and Development Canada
Facilitator

LCDR Leedjia Svec, Ph.D.
NASA Ames Research Center
Organizer

Mr. William E. Tarry, Jr.
Institute for Defense Analyses (IDA)
Participant

Mr. Greg Thornton
US DOT Maritime Administration (MARAD)
Participant

Mr. Brendan Tougher
Anthropocene Institute
Participant

Mr. Pierre-Damien Vaujour
Spire
Participant

Dr. Alonso Vera
NASA Ames Research Center
Speaker

Ms. Wendy Walsh
Naval Postgraduate School
Organizer, Facilitator

Dr. Raechel Waters
Vulcan Inc.
Participant

Mr. Neil Watts
United Nations
Speaker

Dr. Hans Wehn
MDA
Participant, Sponsor

Commander Renzo Whittembury
Peruvian Navy
Participant

Dr. Christopher Wilcox
CSIRO
Participant

Mr. John Wilkins
US Coast Guard
Participant

LT Robert Wilkins, US Navy
ONI
Participant

Mr. Mark Winding
US Coast Guard
Participant

Mrs. Ashley Witthar
Defense Intelligence Agency
Participant

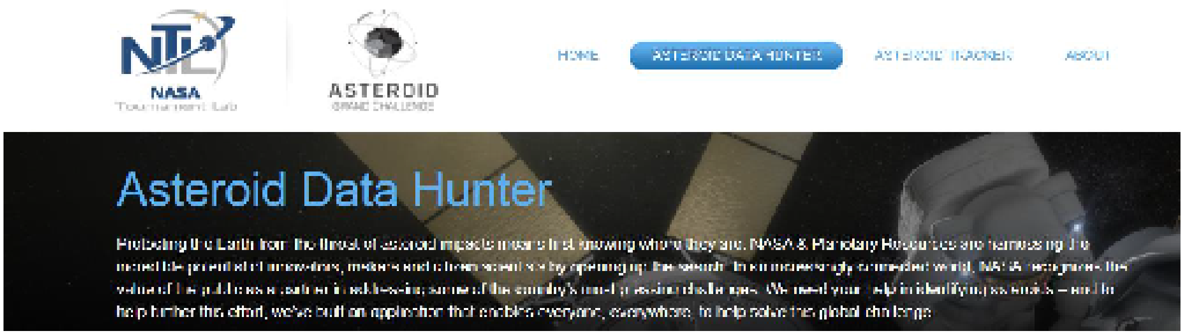
Cdr Joshua Yanchus
Royal Canadian Navy Maritime Forces
Pacific Formation
Participant

Mr. Mark Young
The Pew Charitable Trusts
Participant

Mr. Suzuoki Yukihiro
Japan Space Imaging Corporation
Participant

Supplemental Materials

Case Study 1



NASA's Asteroid Data Hunter Challenge: The Asteroid Data Hunter Challenge, a series of contests with a $55,000 total prize purse, is the first open innovation project to be conducted in support of NASA's Asteroid Grand Challenge. The Asteroid Grand Challenge is a NASA-coordinated global effort to find all asteroid threats to human populations and know what to do about them.

NASA partnered with Planetary Resources, Inc. to conduct this series of challenges leveraging Catalina Sky Survey data hosted on the Amazon cloud. The Asteroid Data Hunter Challenge had two objectives:

> (1) develop a more computationally efficient, general purpose algorithm to detect moving objects using imagery captured by ground-based telescopes—specifically, to increase the detection sensitivity, minimize the number of false positives, ignore imperfections in the data, and ensure that the algorithm can run effectively on all computers; and
>
> (2) develop a software application that is so easy to use that citizen scientists, hobbyist astronomers, and even professional institutions will want to download it. Thus the challenge sought both to provide scientific value to the asteroid detection community and to create a tool allowing citizen scientists to contribute genuinely to asteroid detection.

This challenge was announced to the developer community at the 2014 South by Southwest Festival in Austin, TX, and it received significant press coverage. Throughout the 10 months of the challenge, over 1,200 solvers participated in submitting 700 solutions that resulted in the development of both a new algorithm and a software package. The algorithm has resulted in a 15 percent improvement over the current method of identifying asteroids in the main belt of Asteroids that orbit between Mars and Jupiter.

"The beauty of such archives is that the data doesn't grow stale, and with novel approaches, techniques and algorithms, they can be harvested for new information. The participants of the Asteroid Data Hunter challenge did just that, probing observations of the night sky for new asteroids that might have slipped through the software cracks the first time the images were analyzed," said Jose Luis Galache of the Minor Planet Center. "Moreover, this software can now be used to analyze new images and is available to any observer who wants to use it. The Minor Planet Center applauds these efforts to provide superior tools to all, and looks forward to receiving new asteroid observations generated with them." [1]

The results of this challenge continue to support the use of crowdsourced algorithms to advance NASA's image processing capabilities, as well as the continued use of open innovation in support of the Asteroid Grand Challenge. Furthermore, these results were obtained for a total project cost of less than $200,000, which is less than the salary for one full-time engineer for the same time period.

[1] http://www.nasa.gov/press/2015/march/new-desktop-application-has-potential-to-increase-asteroid-detection-now-available/

GE Tackles the Industrial Internet

Since 2012, General Electric (GE) has pursued a multi-year initiative to leverage the "Industrial Internet" of smart machines and advanced analytics. As part of this vision, GE engaged Kaggle to organize and run three competitions, offering prizes totaling $600,000 to solve high-impact business problems from flight efficiency to hospital operations.

Flight Quest I and II— Changing the future of flight

Flight dynamics change quickly. From weather to gate conflicts, efficiently adapting to changing flight conditions can save millions of dollars in annual fuel costs, as well as reducing carbon emissions. Flight Quest tackled this real-time big data analysis challenge.

In Flight Quest I, participants were given multi-source flight and weather data and asked to predict precise runway and gate arrival times for domestic flights in the United States. The winners produced a 40% accuracy improvement over industry standards—equivalent to saving 5 minutes at the gate per flight (an annual savings of $6.2 million for a mid-sized airline).

Flight Quest II was even more challenging: Participants optimized flights in real time. The second phase included significantly more complex weather data—rain, wind, barometric pressure, ice, and more—as well as crew and passenger counts, airport traffic, and no-fly-zones. The winning solution was evaluated in a flight simulator and found to be a 12% efficiency and cost improvement over real flights.

Hospital Quest— Making hospital visits hassle-free

From long wait times to procedure delays, healthcare inefficiencies cost an estimated $100 billion every year in the United States alone. To help improve patient experience, GE partnered with Kaggle to sponsor Hospital Quest. Participants developed healthcare solutions that were evaluated by experts for quality, impact, and ease of implementation. The winning team implemented a referral management system to simplify hospital discharge workflows and greatly improve the quality of post-acute care.

Further reading about GE Quests—

GE Reports

Minds & Machines keynote

Forbes

O'Reilly

	Flight Quest I	Flight Quest 2
Industry domain	Aviation	
Data Type	Flight history, status, weather, route data	Flight history, status, weather, route and no-fly-zone data
Task	Predict runway & gate arrival time	Choose optimal
Participants	236 players on 173 teams	257 players across 223 teams
No. of entries	3067	3841
Competition length	4 months	6 months
Winning Method	Ensemble of gradient boosting & random forest models	Optimization Techniques
Prize	Total prize pool of $250k	Total prize pool of $250k

The Heritage Health Prize: Bringing Data Science to Preventative Medicine

More than 71 million people are hospitalized every year in the United States, creating at least $30 billion in annual avoidable costs. Thinking big, the Heritage Provider Network (HPN) launched the **Heritage Health Prize** to spur new algorithms that could allow care providers to reach patients before emergencies occur. Such models would revolutionize preventative medicine by reducing unnecessary hospitalizations and improving the overall health of patients. HPN chose Kaggle to run their challenge.

The Heritage Health Prize ran from 2011-2013. Given anonymized claims and provider data, participants were asked to predict which days each patient would spend in the hospital within the next year. Prizes included $500,000 for the final winner and $230,000 in milestone prizes awarded throughout the competition. In addition, HPN reserved a $3MM Grand Prize, contingent on a very high threshold for accuracy. In all, more than 1600 data scientists competed, submitting more than 25,000 models. The participants included two Nobel Prize winners as well as physicians, scientists, actuaries, and others.

Often, there is a tradeoff between data anonymization and predictive accuracy. Some anonymization intentionally causes a loss or blurring of information—giving less to a data scientist from which to train the model. The scrubbing of sensitive patient data for the Heritage Health Prize was complex enough to merit a peer-reviewed journal article. Ultimately with this anonymization, none of the teams reached the accuracy required for the Grand Prize. However, many new approaches were developed through the competition, moving the field forward. HPN proved to be a thought leader by engaging Kaggle's community—showing what is possible with existing health care data and paving the way for radical improvements to come.

Further reading—

Final winners' announcement

Forbes article

Milestone Winners

Quick Facts

Industry domain	Health Care
Data Type	Anonymised health claims and patient data
Task	Predict number of hospitalization days for each patient within the next year
Participants	1,660 participants on 1,353 teams
No. of entries	25,316
Competition length	2 years
Prize	$730,000. Grand prize of $3MM was not awarded.

Abbreviations

ASCM	anti-ship cruise missile
AIS	Automatic Identification System
API	Application Program Interface
ARC	Ames Research Center
AWS	Amazon Web Services
BLOS	beyond line-of-sight
C4I	Command, Control, Communications, Computers, and Intelligence
CEED	Crowd-driven Ecosystem for Evolutionary Design
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CIA	Central Intelligence Agency
CIOP-MDA	Cooperative Interagency Partnership for Maritime Domain Awareness
CMRE	Center for Maritime Research and Experimentation
CMS	Carbon Monitoring System
COA	Certificate of Authorization
COP	Common Operating Picture
CoSSAR	Center for Collaborative Systems for Security, Safety and Regional Resilience
CVISR	Consolidated Vessel Information and Security Reporting
DIC	data intensive computing
DoD	Department of Defense
DPRK	Democratic People's Republic of Korea
EEZ	economic exclusion zone
ELINT	electronic intelligence
ERP	Enterprise Resource Planning
ETL	Extract, Transform and Load
EU JRC	European Commission Joint Research Centre
FAA	Federal Aviation Administration
FINRA	Financial Industry Regulatory Authority
GE	General Electric
GCCS-M	Global Command and Control System - Maritime
GIS	Geographical Information System
GMCC	Global Maritime Operational Threat Response (MOTR) Coordination Center
GMCOI	Global Maritime Community of Interest
GMF	Global Maritime Forum
GPS	Global Positioning System
HD	high definition
HPC	high performance computing
HSC	high scalability computing
IR	infrared
ISR	Intelligence Reconnaissance Surveillance
ITAR	International Traffic in Arms Regulations
IUU	illegal, unreported, and unregulated
LCS	Littoral Combat Ship
LHC	Large Hadron Collider
LRIT	long-range identification and tracking
MAP-21	Moving Ahead for Progress in the 21st Century Act
MARAD	Maritime Administration
MDA	Maritime Domain Awareness or MacDonald, Dettwiler, and Associates
ML	Machine Learning
MOC	Maritime Operational Center
MOTR	Maritime Operation Threat Response
MPA	Marine Protected Area
MSA	Maritime Situational Awareness
MSSIS	Maritime Safety and Security Information System
NAFD	North American Forest Dynamics
NAS	NASA Advanced Supercomputing or National Airspace System
NCA	National Climate Assessment
NEX	NASA Earth Exchange
NGO	non-government organizations
NMIO	National Maritime Intelligence-Integration Office
NOAA	National Oceanic and Atmospheric Administration
OSINT	open source intelligence
PAE	Pacific Architects and Engineers Incorporated
PM-ISE	Program Manager for the Information Sharing Environment
PMAR	Piracy, Maritime Awareness and Risks
PoE	Panel of Expert
R&D	research and development
SAP	Systems, Applications, and Products in Data Processing
SAR	synthetic aperture radar
SCR	Security Council Resolution
SGI	Silicon Graphics International Corp
SKA	Square Kilometer Array
SOCINT	social intelligence
TREAD	Traffic Route Extraction and Anomaly Detection
UAE	United Arab Emirates
UAS	unmanned aerial system
UN	United Nations
USC	University of Southern California
VMS	Vessel Monitoring System
WAM-V	Wave Adaptive Modular Vessel
WELD	Web Enabled Landsat Data
WMD	weapons of mass destruction
ZB	Zettabyte