ACQUISITION INNOVATION RESEARCH CENTER

Defense Data Grand Prix I

Final Technical Report

EXECUTIVE SUMMARY AND REPORT

JUNE 2023

PRINCIPAL INVESTIGATOR:
Dr. Stoney Trent, *Virginia Tech*

CO-PRINCIPAL INVESTIGATOR:
Dr. Hoong Yan See Tao, *Stevens Institute of Technology*

SPONSORS:
MR. MARK E. KRZYSKO, SES, PRINCIPAL DEPUTY FOR ACQUISITION DATA AND ANALYTICS, OFFICE OF THE UNDER SECRETARY OF DEFENSE FOR ACQUISITION AND SUSTAINMENT
DR. LINDSEY SAUL, CHIEF DATA AND ANALYTICS OFFICER, DEFENSE LOGISTICS AGENCY

RESEARCH TEAM

NAME	ORG.	LABOR CATEGORY
Stoney Trent	Virginia Tech	Principal Investigator (PI)
Hoong Yan See Tao	Stevens Institute of Technology	Co-PI
Will Hanley	Virginia Tech	Research Associate

TABLE OF CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

This final technical report summarizes Year 1 of the Defense Data Grand Prix (DDGP). The DDGP was established in 2021 and in line with the Department of Defense's (DoD) Data Strategy aimed at (a) reduce the two largest barriers (access and knowledge of operational problems) to scaled data analytics in the DoD, (b) incentivize innovations and new perspectives to create unanticipated findings, and (c) increase awareness of Defense Acquisition System challenges, decisions, and processes.

The DDGP is an Acquisition Innovation Research Center (AIRC) competition in which faculty-led teams collaborate with government stakeholders to solve real-world problems. In the first iteration of the DDGP, *Grand Prix I*, 15 teams from seven universities competed across three semester-long Heats. The Defense Logistics Agency (DLA) was the primary operational sponsor for *Grand Prix I*. As such, DLA's Chief Data and Analytics Office orchestrated collaborations with 12 problem owners and worked to pilot the governance processes necessary for DDGP competitors to utilize Controlled Unclassified Information (CUI). *Grand Prix I* concluded in January 2023. Some of the largest barriers to getting academic researchers working with CUI include:

(1) access to real-world problem owners;
(2) disparate government data owners and data sharing policies;
(3) CUI handling training;
(4) lack of non-disclosure agreements tailored for faculty and students who are not employees of a contractor; and
(5) lack of an NIST 800-171 compliant information system accessible by CUI-eligible researchers (who are not on contract and do not have Common Access Cards).

The *Grand Prix* included three Heats over an 18-month period (i.e., three university semesters):

- Heat 1: Describe proposed objectives and recommend approaches to data access and analytics.
- Heat 2: Demonstrate scalable access and sharing of real, transformed, or synthetic defense acquisition data.
- Heat 3: Analyze and visualize findings from defense acquisition data.

Through the first two Heats, *Grand Prix I* reduced the first four barriers. In the culminating third Heat, the U.S. Marine Corps (USMC) Aviation Department participated to pilot the use of a CUI-compliant information system hosted by Virginia Tech: the Defense Acquisition Research Collaboration and Innovation Environment (DARCIE), which enabled two teams (from Stevens Institute of Technology and George Mason University) to access USMC-provided readiness data.

The overall results illustrate the promise of the DDGP and the accomplishment of two key objectives: 1. provide data-driven analysis to inform DoD operational and policy decisions related to data and operations; and 2. allow researchers to use real-world data to tackle real-world problems in ways that integrate with academic courses and research seminars. The DDGP website can be found at: https://acqirc.org/defense-data-grand-prix/.

INTRODUCTION

The Defense Data Grand Prix (DDGP) is a novel problem-solving venue, in which government problem owners collaborate with academic teams to solve real-world problems with real-world data. Additionally, the competition affords strategic decision makers with insights into data and manning strategies. The first iteration of the DDGP, *Grand Prix I*, leveraged partnerships with the Defense Logistics Agency (DLA) and U.S. Marine Corps (USMC) as operational sponsors to provide real-world data and problems for the academic teams. The competition aims to:

- **increase awareness of Defense Acquisition System challenges, decisions, and processes;**
- **inform data strategies through real-world tactical problem-solving; and**
- **encourage innovation to improve national security.**

Grand Prix I was a three-heat, 18-month competition in which faculty-led university teams collaborated with Department of Defense (DoD) stakeholders and problem owners for semester-long projects to develop and recommend solutions to real-world problems. In the process, the competition leverages university capabilities while building an informed talent pipeline. Team submissions were ranked by government sponsors based on the impact, acceptability, suitability, and feasibility of their proposed approaches. Awards for each heat ($40 thousand (K) first place, $30K second place, $20K third place, and $10K fourth place) were provided as AIRC sub-contracts through the winning teams' respective universities. Figure 1 highlights the progressive nature of the sequential heats.

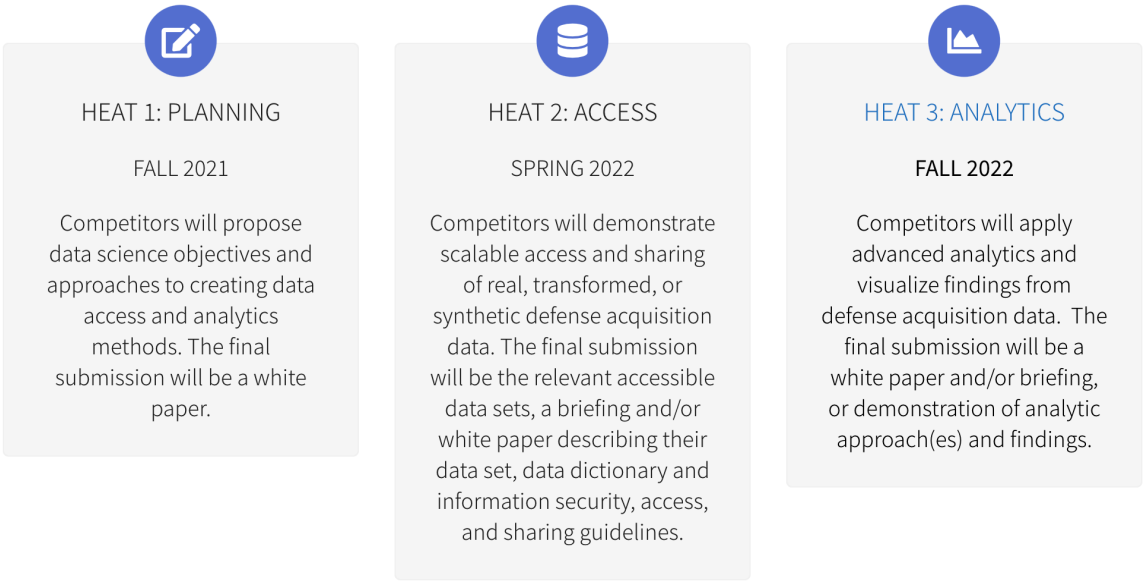


Figure 1. Defense Data Grand Prix Heat Description and Schedule

Grand Prix I presented data science challenges to academic teams that included technical and non-technical barriers to enterprise analytics. For example, the DLA annually procures more than $48 billion in goods and services, including food, equipment, and fuel, on behalf of the U.S. military services and its other customers, generating vast amounts of data in its supply-chain tracking system. DLA makes decisions that must be informed by syntheses of these data. To leverage modern data science approaches, DLA must grapple with system and data architectures and data governance policies. The following sections summarize the submissions from each Heat of *Grand Prix I*.

COMPETITION DESCRIPTION

At the beginning of each heat, competitors were provided an overview of operational problems, data characteristics, and other information relevant for their challenge. This orientation seminar was attended by the Sponsor's problem owners and competing teams via remote video conference, and competitors were allowed to select the top three (3) problems they wanted to work on after this seminar.

Unlike other data competitions in which problems are well defined and sample data are provided, this competition began with problem identification and data inspection. Teams were encouraged to work with government sponsors to understand their problems, data, and

constraints. This collaboration started with an orientation seminar arranged by AIRC at the beginning of each heat. Subsequent interactions between government sponsors and competitors were determined and coordinated by the sponsors. Teams were expected to address ethics, leadership, and project management. Successful teams in this competition would:

1. Participate in government sponsor seminars with other competitors to learn about data, problems and organizations;
2. Document data and information system architectures to determine curation and sharing needs and constraints;
3. Apply ethical and legal considerations;
4. Identify high impact approaches and rationally select among them;
5. Implement mathematical approaches based on sponsor needs;
6. Exercise project management skills and effectively contribute to a team;
7. Interact with a client and deliver the project's outcomes; and
8. Effectively provide or present findings to government sponsors.

Eligibility

The DDGP is open to teams of participants from the universities in the SERC/AIRC collaborator network, Historically Black Colleges and Universities (HBCU) and Minority Institutions (MI). The SERC/AIRC collaborator universities are: Auburn, Carnegie Mellon, Georgetown, Georgia Tech, Massachusetts Institute of Technology (MIT), Old Dominion, Penn State, Purdue, Stevens Institute of Technology (lead), Texas A&M, University of Alabama in Huntsville, University of Massachusetts Amherst, University of Maryland, University of Southern California (USC), University of Virginia, and Virginia Tech. The DDGP is also open to the following universities for this task: George Mason University, North Carolina State University, the University of Arizona, the Ohio State University, and George Washington University. Collaborating military universities include the Air Force Institute of Technology and the Naval Postgraduate School.

Future heats are intended to be open to more institutions, subject to approval by the Contracting Officer. All team members must be from the same academic institution. All members of participating teams must be U.S. citizens or U.S. Permanent Residents and be 18 years of age or older as of 30 August 2021.

Judging

The judging panel consisted of AIRC staff members and representatives from government sponsor agencies. Final award selection was made by DoD officials for each heat.

Each team produced a final product that was evaluated by multiple judges from various partnering institutions and sponsors. Each submission was judged based on the following criteria with each judge independently scoring each solution on a scale of 1–4 for each category. Rankings for each division was selected by adding all judges' scores to determine the highest scoring solution.

Impact: To what degree will the approach positively impact the sponsor's mission?

Acceptability: How broadly can this approach be implemented? Is the approach aligned with DoD and Federal equities?

Suitability: To what degree does the approach suit the needs of the sponsor?

Feasibility: To what degree do technical or workforce hurdles to scaled implementation exist? Are associated costs affordable and commensurate with the expected benefits?

Judging was performed in a "double-blind" fashion to avoid bias. This scheme also required that participants not include any language that may identify them or their previous works (including work done during previous heats or efforts from outside the competition that can be used to identify them).

HEAT 1 SUMMARY (FALL 2021)

In Heat 1, faculty-led teams worked with DLA problem owners to recommend ways to improve access to applicable data. The following were challenge topics that were presented to the teams in Heat 1:

Topic	Operational Problem Description	DLA Organization
Industrial Capability Program Production Input Material Identification	DLA Industrial Capability and Warstopper Program has developed a raw material and components application within the Sustainment Readiness Criteria (SRC) system. Raw material data requirements for DoD spare parts are embedded in various technical data documents. We need a better method to capture the raw material data elements from multiple existing data sources and to identify additional data sources or techniques to capture the data. This information is essential in understanding the risks (e.g., out of production, shortages, sources) associated with the materials/components that feed the production of weapons system and troop sustainment products. Armed with this knowledge the program can propose risk mitigation strategies to help improve product availability.	DLA Acquisition (J7)
Lead-Time Variability Model	Enhance an existing SAS tool which uses historical sampling to model variability in customer demand over procurement lead time for a future period. The goal of this project would be to take the current model, which assumes variable customer demand and a fixed procurement lead time and extend it to consider variable procurement lead time. By improving the current model, DLA can more accurately assess the organic costs required to achieve a target service level when building incentives for performance-based logistics contracts.	DLA Aviation
Operational Sustainment Mapping	Operational Force Sustainment Mapping / Forecasting – using operational force employment and sustainment historical data, combined with anticipated maintenance information (both planned and conditions-based), create a forward-looking sustainment requirements map that results in improved forecasting models/algorithms.	DLA Logistics Operations (J3)
Purchase Request Workload Management Tool	Currently, DLA manages DPC which is a Purchase Request (PR) workload management tool within an Access database. It provides buyers prioritization on what they should work on and history about items that helps them when making their buys. We need to transition from the Access DB to something that is more functional (Access can be clunky) and provides similar or better functionality for our procurement community. The reason it's helpful is because it links a lot of data together for the buying community helping them save time to make their awards.	DLA Land & Maritime (L&M)

The winners of Heat 1 were:

First Place (Dr. Ying Wang and team, Stevens Institute of Technology): The development of a prototype for data visualizations that revealed the bi-directional relationship between products and materials and detected inaccurate descriptions in the current data set in the context of real-time analysis of DLA supply-chain management.

Second Place (Dr. Paul Grogan, Stevens Institute of Technology): The development and demonstration of a new model for Lead-Time Variability that addressed variations in both customer demand and procurement lead times to better estimate the inventory level and government costs required to achieve a target service level.

Third Place (Dr. Christian Lucero and team, Virginia Tech): The completion of a ranked market analysis of modern analytics tools, which could replace the current Microsoft Access-based information processing system for Purchase Request Workload Management. Solutions were proposed based on ease of use, utility, and security.

HEAT 2 SUMMARY (SPRING 2022)

In Heat 2, faculty-led teams worked with DLA problem owners to investigate and recommend ways to implement scalable approaches to data access and sharing of defense acquisition data. The challenge topics in Heat 2 were as follows:

Topic	Operational Problem Description	DLA Organization
AI Technologies, Freedom of Information Act (FOIA) and Operations Security	As required by law, DLA has made procurement data available to the public on the DLA FOIA Library / electronic Reading Room (eRR) website (data set available at https://www.dla.mil/FOIA/Freedom-of-Information/Electronic-Reading-Room/ then scroll down to the section titled "Current DLA Procurement Data Files"). Using advanced data analytics on the publicly available procurement data it is hypothesized that an adversary could potentially derive sensitive information from procurement data. This problem requires competitors to evaluate the publicly available procurement data and determine if that data, either alone, or in combination with other publicly available data sets, would permit an adversary versed in advanced data analytics to derive such information.	DLA Information Governance & Compliance Division (J6DC)
API Based Bidirectional Data Exchange	DLA Aviation needs bidirectional data exchange between EBS and Navy Enterprise Systems via Automated Programming Interface (API) technology to improve timeliness, efficiency, ensuring Zero Trust Architecture, and meeting VAULTIS (Visible, Accessible, Understandable, Linked, Trustworthy, Interoperable, Secure) goals while protecting DLA data and infrastructure.	Office of the Chief Data and Analytics Officer (J6D) & DLA Aviation
Aircraft Downtime Relative to Potential Drivers	Modeling aircraft downtime relative to various potential drivers (e.g., age, use, flight profiles, crews, parts and suppliers). DLA is frequently asked to gauge impact to readiness. Service data is not available to link DLA parts support to Service Readiness.	DLA Aviation
Data Catalog, Data Dictionary, Federated Data Catalogue	Efficient and effective Data Catalog, Data Dictionary, Business Glossary for DLA Enterprise usage and addition to the Federated Data Catalogue. The goal is to identify challenges, propose solution(s), and research best practices specific to DLA's effort to develop a Data Catalog, Data Dictionary, and Business Glossary. It is recommended to develop a high-level proof-of-concept to illustrate the proposed approach(es).	Office of the Chief Data and Analytics Officer (J6D)
Identifying Raw Materials for Industrial Capability Program	DLA Industrial Capability and Warstopper Program has developed a raw material and components application within the Sustainment Readiness Criteria (SRC) system. Raw material data requirements for DoD spare parts are embedded in various technical data documents. We need a better method to capture the raw material data elements from multiple existing data sources and to identify additional data sources or techniques to capture the data. This information is essential in understanding the risks (e.g., out of production, shortages, sources) associated with the materials/ components that feed the production of weapons system and troop sustainment products. Armed with this knowledge the program can propose risk mitigation strategies to help improve product availability.	DLA Acquisition (J7)
Manufacturing Stores and Materiel Shortages	Modeling of Manufacturing Sources and Materiel Shortages. Limited or regulated Obligation Authority, Vendor supply and staffing issues, and raw material scarcity/ shortages present a multitude of manufacturing problems.	DLA Aviation
Stockage Levels and Demands for Consumables and Reparables	Many different planning methods are used to set inventory levels across DLA's network. Aging aircraft and electronic components present supportability challenges due to sporadic demand/diminishing manufacturing sources. DLA needs a tailored approach to level setting.	DLA Aviation

Visible, Accessible, Understandable, Linked, Trustworthy, Interoperable, Secure (VAULTIS) Compliance	Efficient and effective means of becoming a data-centric DLA complying with VAULTIS. The goal is to select three or more of the VAULTIS goals and identify data-related, technological, and organizational challenges, propose solution(s), and research best practices specific to DLA. It is recommended to develop a high-level proof-of-concept to illustrate the approach(es).	Office of the Chief Data and Analytics Officer (J6D)
Zero Trust Architecture (ZTA)	Efficient and effective means and methods of ensuring Zero Trust Architecture (ZTA) throughout DLA. The goal is to explore a Zero Trust Architecture approach that is data-centric and relevant to U.S. government networks, including DLA's, according to concepts and tenets documented in NIST Special Publication (SP) 800-207. It is recommended to also develop diagrams or other visuals to illustrate the proposed approach.	Office of the Chief Data and Analytics Officer (J6D)

The winners of Heat 2 were:

First Place (Dr. Christian Lucero and team, Virginia Tech): The demonstration of approaches to predict supply shortages. The proposed models are flexible and provide several options for DLA Aviation to consider when building their own in-house prediction tool.

Second Place (Dr. Ying Wang and team, Stevens Institute of Technology): The development and demonstration of a proposed architecture to minimize the impact on the existing logistics data system and reduce implementation risks. The resulting augmentation and expansion of data will directly aid the assurance of data quality and positively impact the DLA's mission to identify gaps in the supply chain with a more uniform understanding of the available data.

Third Place (Dr. Roshanak Nilchiani and team, Stevens Institute of Technology): The development and presentation of recommendations to harness the advantages of the joint data catalog and improve its use for artificial intelligence (AI) while increasing security and access/permission control. Recommended technologies offer reduced lifecycle costs while retaining system efficiency, flexibility, robustness, adaptation, and potential for future usability.

Fourth Place (Dr. Benjamin Harvey and team, Bowie State University): The development of a proposed architecture that evaluates the risk associated with releasing unclassified data for Freedom of Information Act (FOIA) requests. The proposed approach includes scraping the dark web for triggering data, performing statistical analysis of the requested data, and then applying privacy algorithms to prevent the release of sensitive information that could threaten U.S. military services.

HEAT 3 SUMMARY (FALL 2022)

Heat 3 kicked off in September 2022 with six universities, eight teams, five DLA problems, and the addition of a new data sponsor, the USMC. Heat 3 focused on analytics and teams applied advanced approaches to derive and visualize findings. The following were the DLA challenge topics for Heat 3:

Topic	Operational Problem Description	DLA Organization
AI/ML Autonomous Vehicle & Robotic Warehousing Research	Autonomous vehicle navigation and autonomous vehicles capabilities, and robotic-assisted warehousing are at different stages of development in the commercial sector. What can DLA learn from these commercial sector advancements that can help make DLA better support our DoD and Whole of Government partners? What are the advantages and disadvantages in the commercial sector implementation of these technologies? What are the risks associated with them? What are DLA's requirements or perquisites needed organizationally and technically to implement these AI/ML technologies?	Chief Data and Analytics Office

DLA Distribution CONUS Storage Utilization	DLA Distribution has over 27 Distribution locations around the world supporting millions of warehouse locations and products. Leveraging industry best practices and tools for Storage and Space Utilization at CONUS Distribution Centers is critical in maintaining a storage and distribution environment that is directly aligned to supporting the warfighter and their needs. DLA Strategic Goal LOE 5.2 is to Assess DLA CONUS facilities footprint to maximize space utilization. This project will be in direct support of this Agency Strategic Goal and will help Distribution HQ understand how best to maximize their storage utilization through modern AI/ML concepts with their data.	DLA Distribution HQ
DLA Records Management Electronic Records Transition	OMB Directive M-19-21 directs all agencies to transition to electronic records this calendar year. DLA offices must implement processes to move to electronic records and the subsequent maintenance, management, and disposition of those records. DLA requires analysis that shows when it makes most sense to scan or maintain hard-copy records (given multiple factors like record lifecycle, number of times accessed, and types of records, etc.). DLA requires a model that shows the financial cross-over point and needed resources for scanning, destruction, and/or long-term storage. What should DLA add to the implementation plan to ensure full agency compliance? What are other agencies implementing that DLA should consider?	Information Governance and Compliance Office; Chief Data and Analytics Office
Identifying Raw Materials for Industrial Capability Program	DLA Industrial Capability and Warstopper Program has developed a raw material and components application within the Sustainment Readiness Criteria (SRC) system. Raw material data requirements for DoD spare parts are embedded in various technical data documents. We need a better method to capture the raw material data elements from multiple existing data sources and to identify additional data sources or techniques to capture the data. This information is essential in understanding the risks (e.g., out of production, shortages, sources) associated with the materials/components that feed the production of weapons system and troop sustainment products. Armed with this knowledge the program can propose risk mitigation strategies to help improve product availability.	DLA Acquisition (J7)
On Time Delivery Predictive Model	DLA has produced a predictive model for time delivery. The predictive model was developed using R, Python, and SAS. It considers about a million rows of 17 predictive variables, which are continuous (cost of order) and categorical (type of order). This model functions well with test and train data. However, its performance suffers when using production data. Can the model's issues be identified? Can a better model be made?	The Analytics Center of Excellence (ACE)

In the DLA Division of Heat 3, the results include:

First Place (Dr. Alireza Zadeh and team, Texas Tech University): The development of a repeatable framework and prediction model to identify late orders with minimal false positives, ultimately reducing a risk for warfighter readiness. The team's model predicted over 96% of all orders and over 83% of late orders were correctly classified. This approach could be broadly implemented across the DLA with computationally cost-effective methods and design.

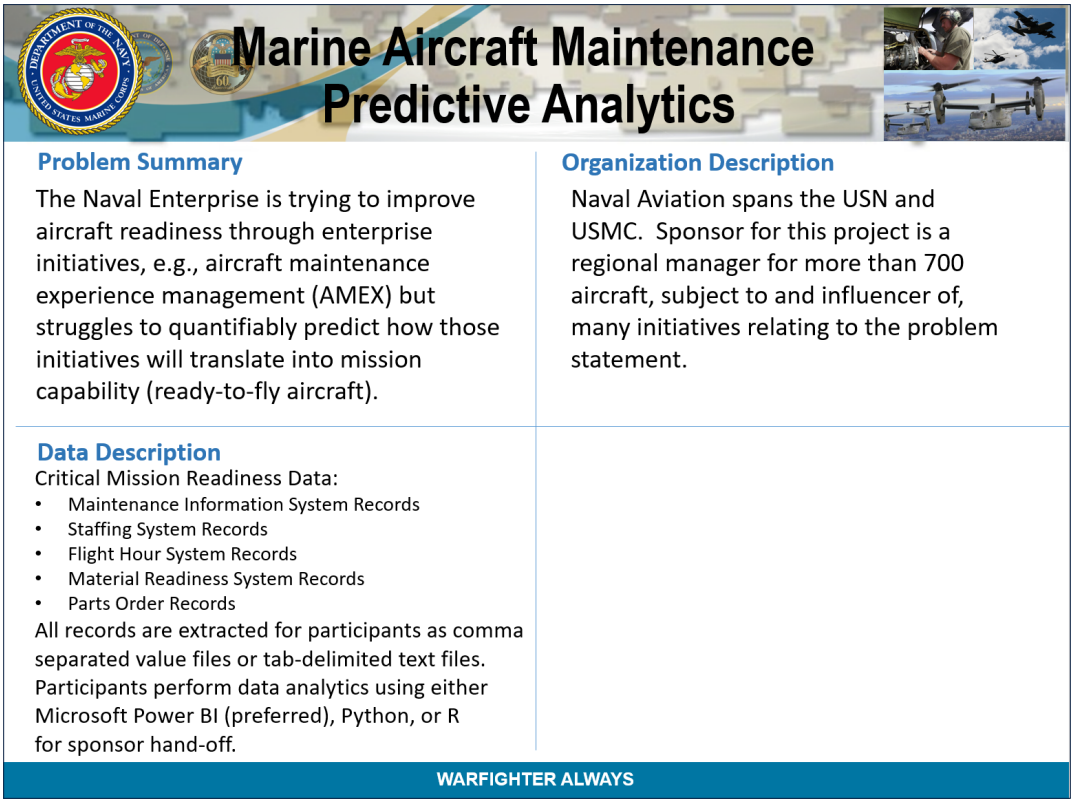
Second Place (Dr. Jacob Hauenstein and team, The University of Alabama in Huntsville): The development of a web-based, interactive model to identify documents to digitize in compliance with OMB Directive M-19-21 as part of the DLA Records Management Electronic Records Transition. This project demonstrated how data science approaches could improve effort allocations in a defense organization.

Third Place (Dr. Feng Liu and team, Stevens Institute of Technology): The development of a holistic analytics platform for warehouse inventory management. The team's descriptive, predictive, prescriptive, and optimization analyses can increase DLA warehouse density by 126% and occupancy score by 28%. If implemented, the adoption of this system, which would require minimal training, could broadly impact DLA operations by significantly improving warehouse utilization.

Fourth Place (Tied) (Dr. Ying Wang and team, Stevens Institute of Technology): The development of a pilot for real-time and long-term risk detection and prediction approaches for clothing and textile supply chains for the Industrial Capability Program.

Fourth Place (Tied) (Dr. Christian Lucero and team, Virginia Tech): The development of a tool to predict on-time delivery of parts. The proposed solution classifies late and on-time orders and performs similarly well on training and testing data.

The USMC challenge topic is described in the following quad chart:



In the USMC Division, competitors piloted the use of a CUI-compliant information system hosted by Virginia Tech. The Defense Acquisition Research Collaboration and Innovation Environment (DARCIE) allows authorized researchers to access DoD acquisition data to derive insights and perform innovative analyses. DARCIE currently runs in a remote desktop environment governed by Virginia Tech's Office of Export Security and Research Compliance (OESRC). DARCIE provides secure, virtual project spaces and access to data science tools such as Python, R Studio, low code dataflow, and Conda package management. DARCIE also provides differential privacy and offers support for advanced analytic techniques, such as machine learning.

In this Division, DARCIE enabled the Stevens Institute of Technology (Stevens) and George Mason University teams to access USMC-provided readiness data, including maintenance records, manning records, part order records, and flight hours. The Stevens and George Mason teams tied, each earning a $25K award for their outstanding efforts. The George Mason team led by Dr. Brett Berlin assessed maintenance efficiency, whereas the Stevens team led by Dr. Ying Wang demonstrated predictive relationships between flight and maintenance hours. Both projects demonstrated how data could more accurately characterize maintenance operations than traditional intuitive approaches. They also demonstrated how readiness data can be used to provide feedback on acquisition decisions such as manning, equipment use profiles, and spare parts.

CONCLUSION

The Defense Data Grand Prix offers teams longer-term exposure to the difficulties found in real-world data science applications. This contrasts with other data competitions that provide discrete and well-defined problems for short-duration demonstrations of analytic approaches. This construct provides teams with an opportunity to engage with problem owners and their data in unique ways. These interactions allow teams to partake in the problem definition as well as the solution development while building relationships that will inevitably contribute to future projects. This structure allows the Grand Prix to run concurrently with strategy revision at multiple echelons in the DoD, allowing the findings from research teams to inform defense acquisition strategy and policy in ways that other competitions do not. Based on the results of Heat 3, both GMU and Stevens winners of the USMC Division submitted a prototype sequel proposal to the USMC to develop a prototype application that aviation maintenance leaders can use to optimize and forecast readiness. AIRC seeks additional DoD data challenges and broader engagement with faculty-led teams from academia for the upcoming *Grand Prix II*.

DISCLAIMER

The Acquisition Innovation Research Center is a multi-university partnership led and managed by the Stevens Institute of Technology and sponsored by the U.S. Department of Defense (DoD) through the Systems Engineering Research Center (SERC)—a DoD University-Affiliated Research Center (UARC).

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)) and the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) under Contract HQ0034-19-D-0003, TO#0480.

The views, findings, conclusions, and recommendations expressed in this material are solely those of the authors and do not necessarily reflect the views or positions of the United States Government (including the Department of Defense (DoD) and any government personnel), Virginia Tech, or Stevens Institute of Technology.

No Warranty.

This Material is furnished on an "as-is" basis. The Stevens Institute of Technology and Virginia Tech make no warranties of any kind—either expressed or implied—as to any matter, including (but not limited to) warranty of fitness for purpose or merchantability, exclusivity, or results obtained from use of the material.

The Stevens Institute of Technology and Virginia Tech do not make any warranty of any kind with respect to freedom from patent, trademark, or copyright infringement.

SYSTEMS ENGINEERING RESEARCH CENTER

AIRC ACQUISITION INNOVATION RESEARCH CENTER

STEVENS INSTITUTE OF TECHNOLOGY 1870