AIRC PERSPECTIVES

JUNE 2022

Defense Data Grand Prix: Solving Grand Problems to Meet a Grand Vision

AIRC
ACQUISITION INNOVATION RESEARCH CENTER

Stoney Trent, Christian Lucero
VIRGINIA TECH

Hoong Yan See Tao, Ying Wang, and **Paul Grogan**
STEVENS INSTITUTE OF TECHNOLOGY

ABSTRACT

The Department of Defense (DoD) seeks ways to engage data-science experts on its compelling, real-world challenges and decisions. In comparison, university students of data science need practical challenges to tackle as senior projects and graduate research. In response, the Acquisition Innovation Research Center (AIRC) created the Defense Data Grand Prix competition to bring these two compelling needs together in a win-win situation. In the Grand Prix, faculty-led university teams collaborate with DoD stakeholders and problem "owners" for semester-long projects to develop and recommend solutions to real-world problems. The competition provides data-driven analysis to inform DoD operational and policy decisions related to data and operations while giving students valuable, real-world data and analytic questions to tackle in ready-made practicums that can be integrated with undergraduate and graduate courses and research seminars. Initial results described in this paper illustrate the promise of this activity. In response, AIRC continues to seek additional DoD data challenges and a broader engagement with faculty-led teams from academia for future "heats" in the Defense Data Grand Prix.

INTRODUCTION

In September 2020, the DoD published a data strategy[1] that supports digital modernization and use of data at the speed and scale needed for operational advantage and increased efficiency. The strategy emphasized the importance of data access, usability, standards, talent, and culture. As a data-intensive function, defense acquisition provides a breadth of strategic and tactical problems to inform DoD decision makers as they fulfil the intent of using data to improve government operations and outcomes.

In parallel, university students of data science need practical challenges to tackle as senior projects and graduate research as part of their education and to give them insight and experience in real-world problems and future employment opportunities.

In response, AIRC created the Defense Data Grand Prix competition to bring these two compelling needs together in a win-win situation. The Defense Data Grand Prix is a three heat, 18-month competition that aims to:

- **Improve data access;**
- **Increase awareness of Defense Acquisition System challenges, decisions, and processes; and**
- **Encourage innovation to generate unanticipated findings to improve the efficiency and effectiveness of national security.**

[1] Deputy Secretary of Defense (DSD), DoD Data Strategy, September 30, 2020b. Cleared for Public Release. https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DoD-Data-Strategy.pdf

Currently in Heat 2 of the first round, the Grand Prix encourages collaboration among academic teams, government sponsors, and corporate partners. It seeks close and frequent interaction between sponsors and competitors, rewards for innovation, and shareable findings. Awards for each heat (currently at $40 thousand (K) first place, $30K second place, $20K third place, and $10K fourth place) are made to the top teams through their respective universities. Submissions are shared amongst all competitors to promulgate best practices and concepts for subsequent heats. Competitor teams may participate in any or all heats. Figure 1 provides further details about each Heat.

HEAT 1: PLANNING

FALL 2021

Competitors will propose data science objectives and approaches to creating data access and analytics methods. The final submission will be a white paper.

HEAT 2: ACCESS

SPRING 2022

Competitors will demonstrate scalable access and sharing of real, transformed, or synthetic defense acquisition data. The final submission will be the relevant accessible data sets, a briefing and/or white paper describing their data set, data dictionary and information security, access, and sharing guidelines.

HEAT 3: ANALYTICS

FALL 2022

Competitors will apply advanced analytics and visualize findings from defense acquisition data. The final submission will be a white paper and/or briefing, or demonstration of analytic approach(es) and findings.

Figure 1 - Defense Data Grand Prix Heat Topics and Schedule

The Defense Data Grand Prix presents challenges to solve the most difficult data science issues facing the DoD today. These include technical and non-technical barriers to enterprise analytics. For example, the Defense Logistics Agency (DLA) annually procures more than $42 billion in goods and services on behalf of the U.S. military services and its other customers including food, equipment, and fuel, generating vast amounts of data in its supply-chain tracking system. DLA—a data sponsor and partner of the Defense Data Grand Prix—makes decisions that must be informed by syntheses of a variety of these data. To leverage modern data science approaches, DLA must grapple with system and data architectures and data governance policies. In Heat 1 of the Defense Data Grand Prix, teams worked with DLA problem owners to recommend ways to improve access to applicable data.

Heat 1 Summary

Heat 1, which concluded at the end of 2021, provided valuable insights and fresh perspectives to DLA through recommendations for data access and usability. Submissions were ranked based on the impact, acceptability, suitability, and feasibility of their proposed approaches. The top three submissions for Heat 1 were:

- First place: **Professor Ying Wang's team from Stevens Institute of Technology proposed data engineering and software architecture to improve material identification for the Industrial Capability Production.**
- Second place: **Professor Paul Grogan from the Stevens Institute of Technology developed a new model for Lead-Time Variability that addressed customer demand and procurement lead times.**
- Third place: **Professor Christian Lucero's team from Virginia Tech provided a comparative assessment of modern replacements for the current, Microsoft Access-based, information processing system for Purchase Request Workload Management.**

Detailed descriptions of the problem areas and results from the top-three submissions are as follows:

1. **First Place: Industrial Capability Program Production Input Material Identification**

   Professor Ying Wang and her team presented initial design and prototypes of a proposed solution to the challenge of "Industrial Capability Program Production Input Material Identification." In this challenge, DLA's Acquisition directorate predicts potential supply-chain shortages and levels of inventories in its distribution centers to maintain adequate circulation of products and materials, setting proper adjustments nationally and globally. It is critical to enable real-time analysis of the logistics information data in the context of DLA supply-chain management to provide dynamic decision support to manage this process.

   The team's proposed solution provided thorough analysis and examinations of the Federal Logistics Information Services (FLIS) data, focusing on the part relevant to clothing and their raw materials. The team presented a prototype for data visualizations that revealed the bidirectional relationship between products and materials, detecting inaccurate descriptions in the current data set. The proposed architecture (see Figure 2) and infrastructure of a comprehensive software stack should enhance the visualizations, transparency, usability, and robustness of DLA's global supply-chain management.

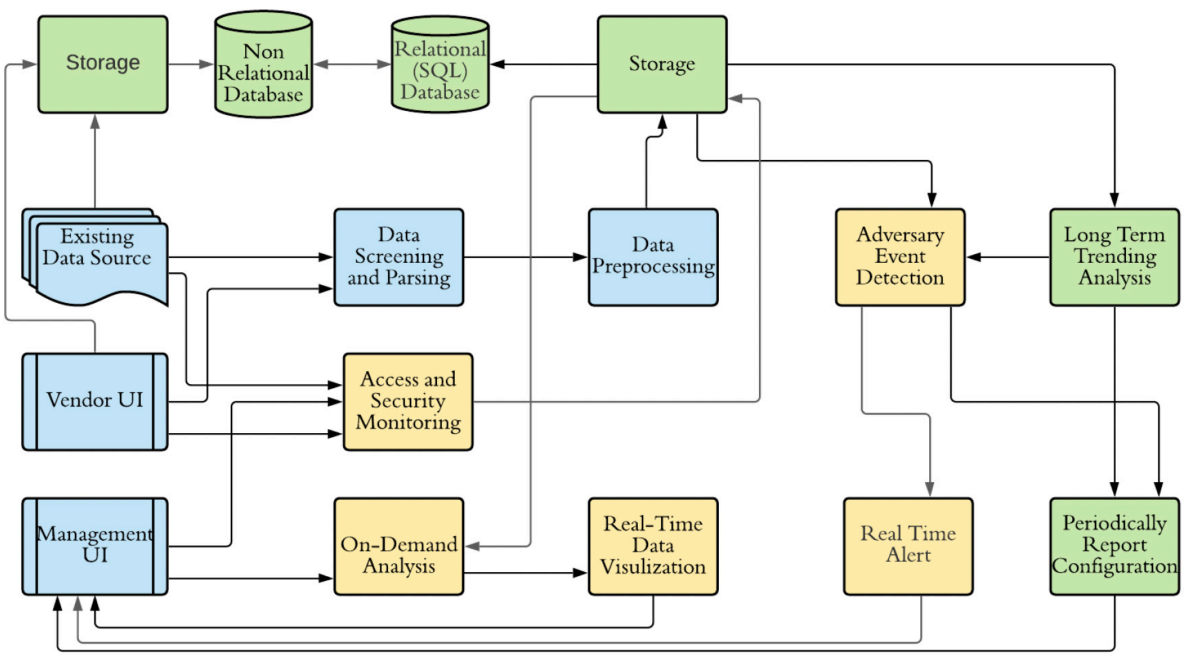


Figure 2 – Overview of the Proposed DLA Supply-Chain System Management Architecture

Based on the analysis of the data set and consultation with the DLA team, the high-level Architectural Significant Requirements (ASRs) were developed and identified as usability, scalability, performance, security, availability, and testability. The development of the ASRs built a critical piece of the system architecture. Given the available data (extracted from a larger DLA data set), data characteristics and relationship were presented and analyzed. The visualizations of the data were designed and demonstrated. After a successful completion in Heat 1, the team advanced to Heat 2 and is currently working on the same project to prepare for the implementations and deployments.

2. **Second Place: Lead-Time Variability Model**

Professor Paul Grogan earned second place in Heat 1, working on the "Lead-Time Variability Model" challenge posed by DLA Aviation. DLA Aviation provides logistics support to its government customers for over 1 million different items that are sourced from producers using standard procurement processes. To guide the timing and size of procurement orders, DLA Aviation aims to maintain an inventory of components to achieve a target service level (i.e., to meet a desired probability of sufficient supply to meet customer demand before the next delivery). This project was motivated by a stated need of DLA Aviation for accurate assessment of the organic (government) costs required to achieve a target service level to help build incentives for performance-based logistics contracts. For example, estimating the government costs to maintain inventory stocks to meet a target service level can define cost targets for supplier contracts to directly provide components to customers (see Figure 3).

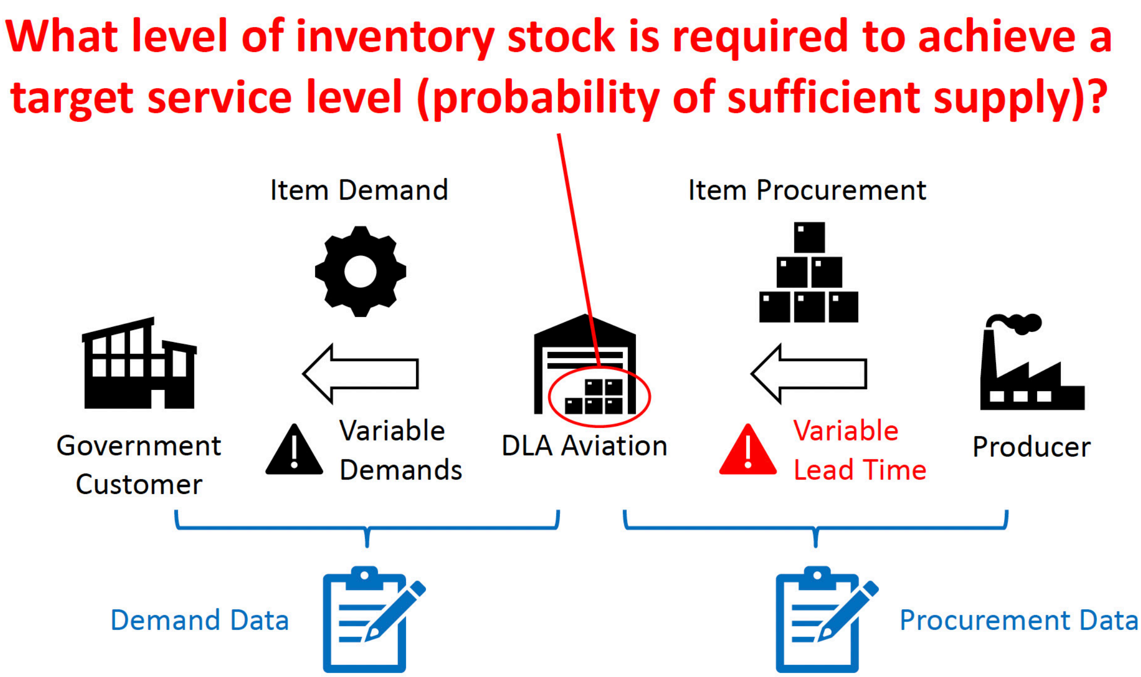


Figure 3 – DLA Aviation Workflow

The current DLA lead-time model considers variable customer demand but fixed procurement lead times, potentially underestimating the required inventory level (and cost) to achieve a target service level. Alternatively, variable procurement lead times should more accurately capture the intrinsic uncertainty in meeting customer demands. The objective of this challenge was to extend the existing DLA model to consider variable lead times to better estimate the inventory required to achieve a target service level.

To demonstrate modern data science methods, Paul Grogan utilized an interactive Jupyter Notebook[2] that combines rich text and Python software code to present the results in a single document (i.e. white paper). Numerical and simulation-based examples use the Python programming language (version 3.8.8) with numpy (v. 1.20.1), scipy (v. 1.6.2), pandas (v. 1.4.1), and matplotlib (v. 3.3.4) libraries (all available under open-source licenses). Some texts were generated from an embedded Python script. All code can be edited and executed in the Jupyter Notebook. The overall modeling approach was presented and demonstrated using fictional data to represent the demand, procurement, and service-level models (see Figure 4).

[2] Jupyter Notebook is a web application for creating and sharing computational documents (https://jupyter.org/).

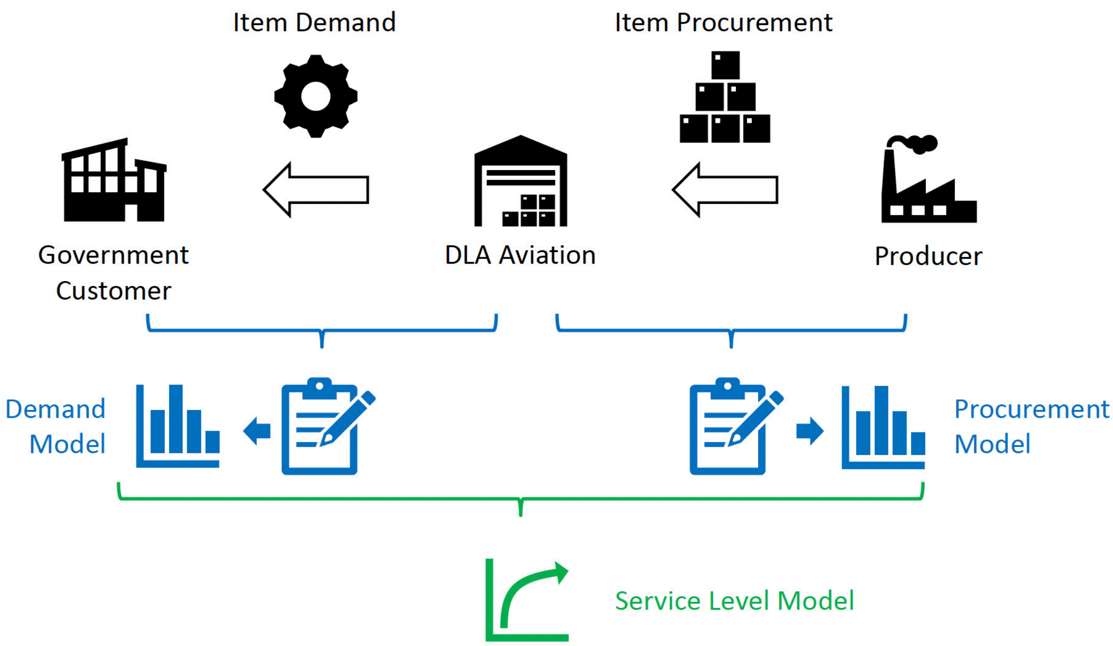


Figure 4 – Model Overview

Major model assumptions of the service-level model and opportunities for future model improvement were generated. Paul Grogan also developed a new model to consider variable customer demand and variable procurement lead times. The results demonstrated that the demand assumptions closely matched the sample data, but more development may be necessary to improve the procurement model.

3. Third Place: Purchase-Request Workload-Management Tool

The third-place team led by Professor Christian Lucero worked on a challenge titled "Purchase-Request Workload-Management Tool." DLA manages over 5 million items supporting over 2,000 different weapon systems. All this is done simultaneously while awarding thousands of contract actions each day. With such high stakes, DLA needs analytics tools that are easy to use, powerful, and secure. DLA Land and Maritime currently uses a Microsoft Access dashboard with SQL and SAS back ends as a custom purchase-request workload management tool. This tool is used by hundreds of employees but is not fully optimized. The team provided a thorough assessment and comparison of modern replacements for the current Microsoft Access-based information-processing system, using a ranking system across various categories, as shown in Table 1.

	Access	Power BI	Power Apps	Qlik	ServiceNow	Tableau	R Shiny
User Interface	1	6	4	5	2	7	3
Dashboard Interactivity	1	6	4	3	2	7	5
Advanced Analytics	1	4	2	6	3	5	7
Ease of Coding	2	3	4	5	6	7	1
Robustness	1	5	2	4	3	6	7
Security	7	7	7	4	3	2	1
Price (user/month)	7	7	7	4	1	2	3
Total Score	**20**	**38**	**30**	**31**	**20**	**36**	**27**

Table 1 – Ranking of Existing Products

The team concluded their product research, which resulted in a ranking of products with Microsoft Power Business Intelligence (BI) winning out and Tableau finishing second.

Future Opportunities

The Defense Data Grand Prix is unlike other data competitions that provide discrete and well-defined problems for short-duration demonstrations of analytic approaches. In contrast, the Defense Data Grand Prix offers teams longer-term exposure to the difficulties found in real-world data science applications. 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 Defense Data Grand Prix to run concurrently with strategy revision at multiple echelons in the DoD, allowing the findings from research teams to inform strategy and policy in ways that other competitions do not.

Heat 2 Is Underway

Six teams are currently collaborating with DLA analysts in Heat 2 to investigate the following problems:

- **Aircraft Downtime Relative to Potential Drivers;**
- **Balancing Freedom of Information with Operations Security;**
- **Data Catalog, Data Dictionary, and Federated Data Catalog**
- **Identifying Input Material for the Industrial Capabilities Program;**
- **Manufacturing Stores and Materiel Shortages; and**
- **Stockage Levels and Demands for Consumables and Reparables.**

The objective for Heat 2 competitors is to demonstrate how to make more and better data accessible for these problems. As with all Heats, the strongest teams will demonstrate not only computer and data science expertise but also policy, system architecture, and system analysis expertise. Submissions are due 30 June 2022.

Registration for Heat 3 begins on 1 July 2022. Register here!

AIRC continues to seek additional DoD data challenges and a broader engagement with faculty-led teams from academia for future "heats" in the Defense Data Grand Prix. Please visit our website **(https://acqirc.org/defense-data-grand-prix/)** for more information or to contact the principal investigators if you have DoD data challenges for Heat 3 or beyond.

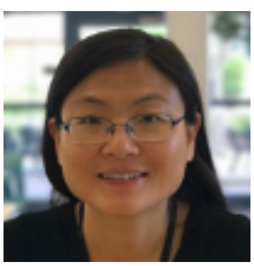
ABOUT THE AUTHORS

Colonel (USA Ret.) Stoney Trent, *Ph.D., AIRC, Virginia Tech*

Dr. Stoney Trent is a Cognitive Engineer and Military Intelligence and Cyber Warfare veteran. Stoney is the principal investigator and lead of the Defense Data Grand Prix. Stoney is also a research professor and principal advisor for research and innovation at Virginia Tech. He is the Founder and President of The Bulls Run Group, a human-machine systems company that designs, delivers, and evaluates technologies for high performance organizations. Specializing in team cognition and human-machine teaming, Dr. Trent designed the Joint Artificial Intelligence Center (JAIC), and established product lines to deliver human-centered AI to improve warfighting and business functions. Previously, he established and directed U.S. Cyber Command's $50M applied research lab, which develops and assesses products for the Cyber Mission Force. Stoney has served as a Strategic Policy Research Fellow with the RAND Arroyo Center and is a former Assistant Professor in the Department of Behavioral Science and Leadership at the United States Military Academy. He has served in combat and stability operations in Iraq, Kosovo, Germany, and Korea. Stoney is a graduate of the Army War College and former Cyber Fellow at the National Security Agency.

Hoong Yan See Tao, Ph.D., *SERC, Stevens Institute of Technology*

Dr. Hoong Yan See Tao is a Research Project Manager for the Systems Engineering Research Center (SERC). Yan is a Co-PI of the Defense Data Grand Prix. She is also a researcher on a current SERC research task titled "DAU Digital Engineering Simulations." Previously, Yan was a SERC Research Associate and Postdoctoral Fellow where she conducted research in mission engineering, digital engineering, and performed project management roles with several human capital development and systems engineering and systems management transformation projects. Yan graduated with a Ph.D. in Systems Engineering from Stevens Institute of Technology in 2017. Yan was the President of the International Council on Systems Engineering (INCOSE) Student Division at Stevens in 2013. Yan received her B.S. and M.S. in Electrical Engineering from the University of Texas at El Paso (UTEP). At UTEP, she worked in the Research Institute for Manufacturing and Engineering Systems and was the team leader for two NASA and Texas Space Grant Consortium design challenges. Yan was also the President of the Society of Women Engineers (SWE) and Institute of Electrical and Electronics Engineers (IEEE) student organizations at UTEP.



Ying Wang, Ph.D., *School of Systems and Enterprises, Stevens Institute of Technology*

Dr. Ying Wang is an associate professor at School of System and Enterprises at Stevens Institute of Technology. She was lead engineer at Commonwealth Cyber Initiative (CCI), Virginia Tech before joining Stevens. She has broad experience in industry and academia, and interdisciplinary knowledge in wireless communications, health informatics, software engineering, and data mining. She was a Distinguished Member of Technical Staff (DMTS) at Verizon and a Senior Wireless Design Engineer at Apple. She graduated at Virginia Tech with her Ph.D. degree focusing on Dynamic Cellular Cognitive System. Her primary research interests including cyber-security, IoT, software architecture, NextG, telemedicine, and edge-centered medical decision support system.

Paul Grogan, Ph.D., *School of Systems and Enterprises, Stevens Institute of Technology*

Dr. Paul Grogan is an assistant professor at the School of Systems and Enterprises at Stevens Institute of Technology. Dr. Grogan directs the Collective Design (CoDe) Lab which studies how engineers design systems with distributed authority, resulting in collective decision-making problems that cannot be solved with traditional systems engineering methods. The research builds on theory and tools in design science, economics, and information science to understand and improve engineering design processes. Research methods develop and use models, simulations, and games to understand how individuals and teams interact with design problems.

Domains of interest include Earth-observing and communications space systems and (terrestrial) critical infrastructure systems including power, water, communications, and transportation. Paul graduated with his Ph.D. from Massachusetts Institute of Technology in Engineering Systems in 2014.

Christian Lucero, Ph.D., ***Statistics and CMDA, Virginia Tech***

Dr. Christian Lucero is an assistant professor of Statistics at Virginia Tech. He is also a core faculty member in the Computational Modeling and Data Analytics (CMDA) program. Christian has a diverse academic background in areas such as Physics, Mathematics, Operations Research, Statistics, and Computer Science and he tries to incorporate all these areas in his teaching. Christian graduated with a Ph.D. in Mathematical and Computer Sciences at the Colorado School of Mines in 2013. His current passions involve experiential learning in data science and statistical computing.

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

This material reflects informed perspectives and ideas to initiate and inform discussions on current topics of DoD interest. The author was supported, in whole or in part, by the DoD 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.