

**DESIGNING THE SCIENTIFIC DATA SUPPLY CHAIN
TO INCREASE KNOWLEDGE CREATION
AND ACCELERATE SCIENTIFIC DISCOVERY
VIA A MULTI-ARMED NEWSVENDOR MODEL**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Nettie H. Roozeboom
August 2023

Copyright © 2023 by Nettie H. Roozeboom
All rights reserved.

ACKNOWLEDGEMENTS

Thank you to Dr. Jim Ostrowski for serving as my advisor and your notable systems thinking. Thank you for your time, support, encouragement, and inclusiveness. Thank you to my committee members, Dr. Wendy Tate, Dr. Andrew Yu, and Dr. Mingzhou Jin. Thank you, Dr. Tate, for emphasizing the difference and important distinction between not-for-profit and for-profit operations. Thank you to Dr. Denise Jackson. Your engineering management theory course was a highlight of my academic course work. Thank you to Ms. Faezeh Bagheri for your friendship and academic collaboration. I look forward to seeing your impact on the world. Lastly, thank you to my family for your support. To my children, I dedicate this to you.

ABSTRACT

The rapid pace of technological advancement has resulted in a growth in data quantities and velocities across all industries, most notably in scientific research. Scientific user facilities owned and managed by the United States federal government are not-for-profit, aim to accelerate scientific discovery and increase knowledge creation by providing a system that supports users to solve difficult scientific issues, and have struggled with the challenge of how to fully use the promise of big data produced by a scientific experimental facility. Inability in the face of this adversity has substantially constrained the scientific research community's capacity for breakthroughs in science. To tackle this challenge, we proposed a new hybrid model called the multi-armed newsvendor model and designed an integrated framework based on it with the intent of fostering scientific discovery. The multi-period newsvendor model emphasizes the importance of information sharing, while the multi-armed bandit framework instructs the customer on how to best address the exploration vs exploitation conundrum. Taking use of the synergy between the two models permits shorter exploratory runs and provides motivation for digitizing the data supply chain to promote near-real-time information sharing. Simulations on the proposed framework manifested by sharing information in real-time with a larger group of domain experts, the customer will maximize yields during a test campaign. Additionally, investigating a case study at NASA within the context of suggested treatment confirmed the striking financial savings in this area.

The aerospace industry is poised to reap the benefits of the data deluge. Specifically in the design phase. The field is rich in physics-based models, and now rich in data. By viewing how the scientific data supply chain is currently arranged and application of the multi-armed newsvendor model, the benefits of information sharing give us reason to digitally transform the scientific data supply chain by changing the transportation method.

The application of the multi-armed newsvendor model will see a broad impact in the areas of scientific research, as well as facility and engineering management. By innovating the transportation method of the scientific data supply chain, technical inclusion and technical accessibility will be enabled.

TABLE OF CONTENTS

CHAPTER ONE INTRODUCTION.....	1
CHAPTER TWO DESIGNING THE SCIENTIFIC DATA SUPPLY CHAIN TO INCREASE KNOWLEDGE CREATION AND ACCELERATE SCIENTIFIC DISCOVERY VIA A MULTI- ARMED NEWSVENDOR MODEL	3
Introduction.....	3
Literature.....	9
Methodology and Solution Approach	14
<i>Multi-Armed Newsvendor Model and Solution Strategy</i>	14
<i>Theoretical Model for Knowledge Creation</i>	18
<i>Scientific Data Supply Chain</i>	23
Discussion	31
<i>Potential Benefit for Scientific Research</i>	31
<i>Potential Benefit for Facility and Operations Management</i>	33
<i>Potential Benefit for Engineering Management</i>	34
Conclusion	35
CHAPTER THREE GROUND TESTING 3.0: EMPOWERED TO MAKE DECISIONS	37
Introduction.....	37
Background	40
Methodology and Solution Approach to Transform the Scientific Data Supply Chain.....	52
Use Case.....	58
Impact	61
<i>Scientific Research</i>	61
<i>Facility and Operations Management</i>	62
<i>Engineering Management</i>	62
Conclusion	65
CHAPTER FOUR CONCLUSION	66
REFERENCES.....	68
VITA	82

LIST OF TABLES

Table 1: Notation and Description.....	19
Table 2: Problem Formulation.....	19
Table 3: Data Rates Within Large, Government-Owned Scientific User Facilities	45

LIST OF FIGURES

Figure 1: Supply Chain Hierarchy in Terms of the Type of Material	5
Figure 2: Physical and Support Supply Chain, Inspired by Carters, Rogers, & Choi (2015)	7
Figure 3: Multi-Armed Newsvendor Model for Scientific User Facility	16
Figure 4: Simulation Logic Flowchart	21
Figure 5: Multi-Armed Newsvendor Simulation Results	22
Figure 6: Traditional Scientific Data Supply Chain	26
Figure 7: Manufacturer's and Customer's Viable Upstream and Downstream Horizon	26
Figure 8: Transformational Scientific Data Supply Chain	28
Figure 9: Data Information Knowledge Wisdom (DIKW) Framework	53
Figure 10: Traditional Scientific Data Supply Chain	54
Figure 11: Transformational Scientific Data Supply Chain	57

CHAPTER ONE

INTRODUCTION

Data volumes are increasing in all industries given the ‘smart’ devices, digitization, and digitalization of the world around us. Each industry must reflect on its goals given this increase in data. The goal for the scientific research community is increase the knowledge created and accelerate scientific discovery. As the core of scientific discovery is data. Science advances by generating and analyzing data that allows us to replace old models with new models that better explain our observations. NASA’s James Webb Space Telescope collects approximately 60 GB of data every day, producing 20 TB of data per year (Gray, 2022). The booming industry of autonomous vehicles research and deployment generates approximately 40 TB of data per eight hours of operations (Waymo, 2023). In medical research, scientists and engineers at Department of Energy and University of Chicago successfully scanned an entire mouse brain using an electronic microscope, producing 1 PB of data per experiment (Foxley, et al., 2021). Each of these examples are on the frontier of science, technology, and engineering, and of those frontiers is exploring phenomena in space and time, which requires a lot of data over both dimensions. With the increase in data available, the question the scientific research community must ask itself is how do we harness the power of big data to increase knowledge creation and accelerate scientific discovery?

To harness the power of big data in scientific research, we need a better understanding of how knowledge is created. In a supply chain, we are always trying to maintain an uninterrupted flow of goods and services. By viewing data as the commodity in a supply chain from the manufacturer to the customer, we can realize a way to improve the process and gain efficiencies. For the scientific data supply chain, the manufacturer, the scientific user facility, wants the customer to extract as much knowledge as possible while testing at the facility. The customer desires to avoid excessive overstock and must avoid understock.

The large, premier scientific user facilities owned and maintained by the United States federal government have incomparable capabilities and sophisticated instrumentation to explore the unknown in the fields of science. These facilities are used to test precious resources, like new fuels, animal subjects, engineered materials, or detailed test articles. The consumption of these resources must be optimized to increase the knowledge created per dollar spent. New and optical technologies combined with automated data acquisition are driving the generation of large datasets, on the order of terabytes (TB) and petabytes (PB). With state-of-the-art instrumentation

available, we are able to generate data at spectacular rates, up to twenty-five gigabytes (GB) per second. At the Department of Energy Argonne National Lab's Combustion Synthesis Research Facility, using state-of-the-art X-ray scattering, a single measurement can be generated in less than a second. Therefore, the ability to determine the next data to collect at a requisite rate is indispensable (Noack & Zwart, 2019), enabling better use of the facility and producing the data the customer actually desires while testing. Given the similarities shared amongst these facilities, this example is true and present at all the large scientific user facilities.

The facility itself is also a precious resource to be used efficiently. The cost to construct these facilities was on the order of hundreds of millions to billions of dollars, and now required tens of millions of dollars to maintain their operation annually. However, the data inventory also presents a cost, which is often overlooked or hidden. For the manufacturer, the data often becomes obsolete, is lost, or fails to meet the customer's requirements. By practice, the scientific user facility, the manufacturer, will contract with the customer to store the data for a finite amount of time to avoid excessive cost and risks. The customer faces the same risk as the manufacturer with the cost to store data inventory.

While the customers to these facilities are from the public and private sectors, the goal of these not-for-profit scientific research facilities is not profit driven, but rather, the goal is to accelerate scientific discovery and increase knowledge creation by providing a system that allows customers to efficiently solve their complex scientific challenges.

For these tremendously large datasets, improved and requisite computational resources must be brought to bear to process the raw data to a new form of usefulness to the customer. We know from our own personal experiences, information devalues with time and the largest value is realized when shared quickly, efficiently, and among a technically diverse set of customers. The same is true for scientific data. In addition, in order to do this and maintain the continuous flow of data from manufacturing to consumption, an improved transportation method must be employed. It is the value of information sharing that gives motivation to digitally transform the scientific data supply chain that will increase knowledge creation and accelerate scientific discovery.

CHAPTER TWO

DESIGNING THE SCIENTIFIC DATA SUPPLY CHAIN TO INCREASE KNOWLEDGE CREATION AND ACCELERATE SCIENTIFIC DISCOVERY VIA A MULTI-ARMED NEWSVENDOR MODEL

Introduction

The process of gaining new information or insights via scientific study is referred to as scientific discovery. It entails using the scientific method, which is a methodical way to research natural occurrences. Scientific discovery could take numerous forms, from the creation of new ideas or models to the finding of new phenomena or the invention of new technologies (Schickore, 2014; Yang, Hu, & Zhou, 2021). Examples of scientific breakthroughs include the identification of the DNA structure (Byrne & Miescher, 2019), the advancement of the theory of relativity (Crispino & Kennefick, 2019), the finding of the Higgs boson particle (De Blas, et al., 2020), and the establishment of the internet (De Blas, et al., 2020). These findings have had a significant influence on our comprehension of the world and our capacity to address complicated issues, and they have sparked innumerable developments in industries like engineering (Lee, 2022), technology (Rodgers, Hay, & Lussos, 2022), and medicine (Fortunato, et al., 2018).

At the core of scientific discovery is the availability of data. Gathering and analyzing data is crucial for creating and testing ideas, developing models, and gaining insights that might lead to novel discoveries (Fortunato, et al., 2018). With the advent of Industry 4.0 (Rüßmann, et al., 2015), businesses in every sector must reevaluate their mission considering the plethora of new data at their disposal. Data volumes, the amount of data generated by various sources, are rapidly increasing with the proliferation of digital technologies and the Internet of Things (IoT) as the result of Industry 4.0 revolution (Peter, Pradhan, & Mbohwa, 2017). This data could open new avenues for development and growth for businesses, but it also poses several new obstacles. On one hand, data analytics could aid businesses in enhancing the quality of their choices, the satisfaction of their customers, the effectiveness of their operations, and the discovery of novel business models. Large data sets, however, could be difficult to handle and store, necessitating both robust infrastructure and appropriate data management procedures (Richins, Stapleton, Stratopoulos, & Wong, 2017). Companies must strike a balance between the benefits and drawbacks of using data to its fullest potential as data volumes continue to rise (Ghasemaghaei & Calic, 2019). In the context of scientific research, the added value of big data is a generation of new knowledge, but still, the challenges of exploiting it persist (Shi, 2022). In the modern digital landscape, data has become a

fundamental building block for all businesses to construct their products and services. The ability to utilize data effectively has become a key driver of success in the dynamic business environment of today (Chen, Preston, & Swink, 2015). Having a clear understanding of how data plays a role in establishing a competitive advantage has become paramount.

Data, like any other commodity, needs careful planning and coordination across all participants across the chain to ensure quality and timely delivery. Inspired by the idea from the management area, Spanaki et al. (2018) designed the Supply Chain (SC) architecture for the data domain to tackle the difficulties of data exploitation. There are many articles about big data in supply chain management (Richey, Morgan, Lindsey-Hall, & Adams, 2016; Brinch, 2018), but Spanaki et al. (2018) distinguishes between “data about the SC” and “data that are the main artifact moving through the SC”.

Through scrutinizing the application of the supply chain, we reached out to Figure 1 As shown, the material supply chain is the general form of the supply chain that could be categorized into different types in terms of the final product. Rafique et al. (2017) introduced the Energy SC and motivated unique features which distinguished energy from other material supply chains, such as the artifact (energy) is difficult to store in large volumes and is lost when transmitted. The Data Supply Chain (DSC) is another example of a unique supply chain that moves a physical good but has only begun to receive attention in recent years following the Industry 4.0 revolution (2018).

The DSC processing maintains both the quality and quantity of data and does not result in any loss or disappearance of data. Unlike physical raw materials in an SC, data can act as inputs, intermediaries, and final products. Since data has the potential to generate significant value, DSC which the major focus of all subcategories of DSC is narrowed into a specific application can be a continuous process (El Kadiri, et al., 2016). In contrast to scientific DSC (SDSC), for including: healthcare (Dirks & Stepro, 2023), weather, financial (Bhatnagar, 2010), and air traffic management (Keller, Ranjan, Wei, & Eshow, 2016), the utilization is different. The SDSC which focuses on managing scientific data to facilitate research and collaboration among scientists could be involved in other SC and created added value for them. The term "scientific data supply chain" could be interpreted as a paradigm for the management, distribution, and sale of processed scientific data beginning with the acquisition of raw scientific data. When the scientific data produced during an experimental test campaign is treated as a product to be managed, the SC will benefit from the same cost-savings as are realized in the material SC. Integrating all sources of

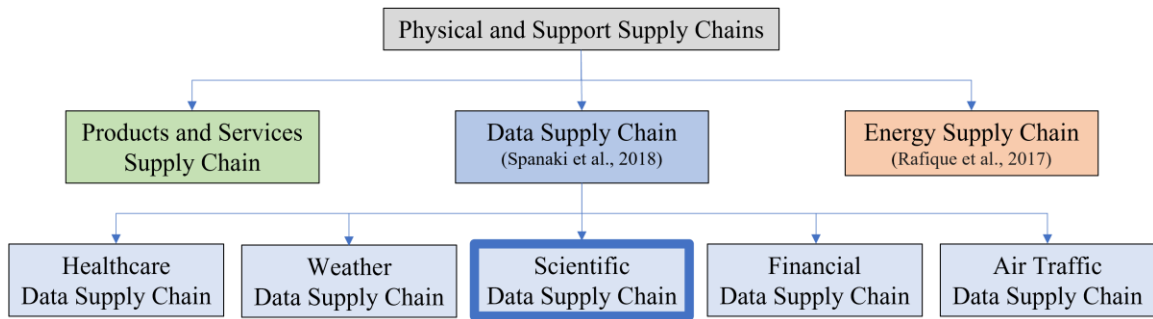


Figure 1: Supply Chain Hierarchy in Terms of the Type of Material

scientific data will provide far-reaching effects (Parkinson, Krishnan, Ushizima, Henderson, & Cholia, 2020).

Through the lens of complex adaptive systems, Carter, Rogers, and Choi (2015) articulated six underlying premises about a supply chain's structure and boundary. The SDSC complies with all six of the stated requirements. While the SDSC can be reduced to a very simple diagram, as shown in Figure 2, it's like many other SCs, "dynamic, complex, and difficult to predict and control" and must be managed as a system (Frankel, Bolumole, Eltantawy, Paulraj, & Gundlach, 2008).

Each supply chain is tailored to achieve a certain result by adjusting and manipulating its underlying structure. Berenguer et al. (2017) argued that while there are several similarities between non-for-profit and for-profit operations in terms of capacitated resource planning, the features parallel with non-for-profit operations are ultimately different from for-profit operations.

As non-profit organizations, the premier scientific user facilities, owned by the U.S. federal government and managed by federal departments or agencies, such as the Department of Energy, the Department of Defense, or the National Aeronautics and Space Administration (NASA), are not in business to make a financial profit. In fact, due to the societal merit and scientific value produced by these facilities, the U.S. federal government prioritizes allocating funding for ensuring they continue operating.

The goal of these scientific user facilities as not-for-profit entities is to support their social values and maximize societal good (Tereyağoglu, Fader, & Veeraraghavan, 2017). These facilities are considered a social business and follow a business model that does not try to maximize profit, like most businesses. Instead, they focus on the "most pressing societal needs" (Bals & Tate, 2017) and for the scientific community, the pressing need is having a state-of-the-art user facility made available to the private and public sector to increase knowledge creation and accelerate scientific discovery. The costs of the sophisticated facilities, technology, and workforce are far too great for a single private company to operate independently. Therefore, the use of these facilities is subsidized by the federal government to decrease the costs to the private and public sectors, however, revenue is collected by charging a fee to each customer. By charging a fee, the overall budget of the facility increases allowing for higher quality services and capabilities to be delivered while also increasing customer participation (Berenguer, Keskinocak, Shanthikumar, Swaminathan, & Van Wassenhove, 2017)

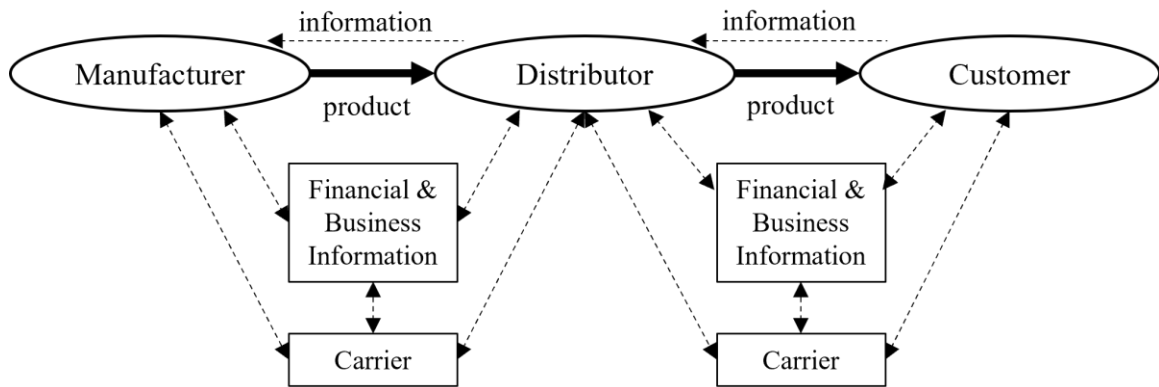


Figure 2: Physical and Support Supply Chain, Inspired by Carters, Rogers, & Choi (2015)

In these scientific user facilities, there is access to state-of-the-art technology that enables the execution of increasingly sophisticated experimental test campaigns by collecting data at high spatiotemporal resolutions to advance the scientific frontier. These high spatiotemporal measurements are often optical and capable of producing data at unprecedented volumes and velocities for these scientific user facilities (Stremper, Yoshii, Hammer, Bycul, & Miceli, 2021). The local computation and storage needs of scientific experimental communities have grown beyond what is available (Bear, Parkinson, Gerhardt, Sibony, & Enders, 2021) to the point that high-performance computing facilities are not just desired by experimental communities but required.

Aryal et al. (2018) claim that 50,000 gigabytes (GB) of data are collected every second compared to 100 GB a day 25 years ago. The NASA scientists, Cox and Ellsworth, (1997) were the first to document this growing concern regarding challenges to manipulating large data sets and the proper computer hardware to manipulate it, called big data. The experimental data generated at scientific user facilities fit the literature's definition of big data (Baaziz & Quoniam, 2015). Keller et al. (2016) highlight that, if data can be integrated, queried, and analyzed, it can allow new discoveries that can impact operational efficiency, and operational safety, and enable strategic decision-making.

According to a study by Vescovi, et al. (2020), scientific user facilities have not systematically harnessed the power of big data sets. So, they are unable to either improve their grasp of how their own facility operates or boost knowledge creation for their customers. The Department of Energy's 2013 Accelerating Scientific Knowledge Discovery (ASKD) Working Group Report (Department of Energy, 2013) stated that a "new advanced discovery ecosystem" could be a promising solution for tackling this weakness.

With the rise in available data, scientific user facilities are entering a new phase in which customers and operators of facilities are empowered to make decisions (Davenport & Kirby, 2015). The time and effort needed to get insight into a larger data collection are proportional to the number of resources dedicated to processing the data. Delaying access to scientific experimental data devalues the information with time (He, et al., 2006; Tran-Thanh, Rogers, & Jennings, 2012).

Through empirical observation while managing tests and deploying state-of-the-art technology within one of NASA's flagship scientific user facilities, we comprehended that the absence of strong infrastructure is not the only factor contributing to the inadequacy of scientific user facilities in terms of scientific discovery and knowledge generation. This setting indeed presents several difficulties that have hampered its use. First of all, scientific user facilities are embedded systems.

This setting is not very versatile, difficult to work with, and costly to develop (Ebert & Duarte, 2018). Secondly, their configuration limits communication with the outside world to an individual's resources, not enterprise-level resources, making it difficult to collaborate with other scientists (Lazarova-Molnar, Mohamed, & Al-Jaroodi, 2018). Finally, the number of subject matter experts who can participate in a test campaign is constrained by the physical size of the facility or the financial resources available to the organization. This could result in reducing the quality of testing and innovation, increasing costs, and delaying testing (Cooke & Hilton, 2015).

The challenges mentioned above, along with the large amount of data that can only be accessed while testing at a specific facility, have led to a slower-than-expected pace of knowledge creation. To address this issue, by relying on analogical reasoning from the area of SC management, we suggested a novel hybrid model that we refer to as the multi-armed newsvendor model and built an integrated framework around it intending to foster scientific discovery. Although the multi-period newsvendor model (Qin, Wang, Vakharia, Chen, & Seref, 2011) highlights the value of information exchange, the multi-armed bandit (MAB) framework (Slivkins, 2019) teaches the consumer how to optimally handle the exploration vs. exploitation conundrum. Using the complementary nature of the two models allows for faster iteration of hypotheses and serves as the impetus for digitizing the data supply chain to facilitate near-real-time information exchange.

The rest of this paper is organized as follows. Section 2 provides background information on supply chain management, which is useful for innovating the scientific data supply chain. Section 3 proposes a new hybrid model, the multi-armed newsvendor model, as a framework for customers when testing at large scientific user facilities. The model is designed to help customers understand complex phenomena where exploration is extremely expensive and must heuristically make decisions based on the data as it is shared. Section 4 gives a retrospective use case at NASA where a major NASA program that would have benefitted from the multi-armed newsvendor model framework to explore effectively if the scientific data supply chain had been digitally transformed prior to testing. Section 5 shares practical implications for the multi-armed newsvendor model and the impact on facility, program, and engineering managers. Section 6 concludes the research.

Literature

In this study, the challenges of scientific user facilities in terms of their impotence in exploiting the full power of data for knowledge creation, motivate us to seek a resolution for tackling. To this end, we take the advantage of analogical reasoning—taking something familiar and making it new or vice versa (Gentner & Smith, 2012). Before the concept of gravity and physical forces existed, German astronomer Johannes Kepler, credited as the inventor of astrophysics, used a series of

analogies from heat, odor, light, magnetism, fluids, and optics to propose that planets travel elliptical paths around the sun based on their distance (Gentner, et al., 1997). Epstein (2019) highlights that the genetic algorithms used in the field of optimization are inspired by Charles Darwin's theory of natural selection. Evidently, the use of analogical reasoning goes beyond the scope of this particular context. In the present study, analogous reasoning serves as a cornerstone in the investigation of how scientific user facilities could hasten the pace of discovery.

Scientific user facilities are continuously trying to strike a balance between two key objectives: making the most of the data that is already accessible and deciding what additional data will be required for a more in-depth study. Admitting doing more trials yields more important data, funding and resource constraints restrict scientists from doing so. The resolution to this dilemma is equivalent to the answer of the "exploration vs exploitation" conundrum.

In the literature, the best method to approach this conundrum is the multi-armed bandit (MAB) paradigm. The MAB framework, formalized by (Robbins, 1952), is useful for sequential decision-making problems in which a limited number of resources must be allocated between competing choices. Each of these choices has an unknown reward distribution, and the user wants to produce the highest possible reward. Often, the number of chances to "pull the arm" is limited by time or budget. Therefore, the MAB's intuitive framework, robust performance, and ability to learn from less feedback (Bouneffouf, Rish, & Cecchi, 2017) produce a powerful framework for business operations (Schwartz, Bradlow, & Fader, 2017). The user is required to first investigate all available choices in order to comprehend the reward system. After an acceptable solution has been found, the user has the choice to make use of that solution.

Several tools in the MAB toolbox give users flexibility for a wide range of use cases (Slivkins, 2019). Thompson Sampling (Thompson, 1933) the family of upper confidence bound algorithms (Lai, 1987), epsilon-greedy, and Boltzmann exploration/SoftMax (Sutton & Barto, 2015), reinforcement learning (Zhang & Bareinboim, 2017; Auer, Cesa-Bianchi, & Fischer, 2002) are a few of the popular algorithms that have been developed through the idea of MAB.

The underlying theme within the current literature is that the choice of a MAB model is usually domain-specific, and with each application, a slight variation is applied to the model to maximize its respective reward function. The performance and selection of a MAB algorithm typically depend on the number of arms, the number of trials, and the reward distribution variance (Kuleshova & Precup, 2000).

In the domain of MAB, there are uses for the generalized MAB framework outside designing algorithms. Many fields have benefited from the original idea of it in their research studies. Within

the context of healthcare and clinical trials, (Bouneffouf, Rish, & Cecchi, 2017) investigated how reward processing is altered in individuals with mental disorders. Durand et al. (2018) developed an adaptive and personalized treatment plan for de novo carcinogenesis in mice, with the aim of finding the best treatment for each mouse, while minimizing the number of mice that die.

The MAB in financial portfolio selection dynamically allocates investments to different assets in the portfolio over time to maximize the expected returns while minimizing the risk (Shen, Wang, Jiang, & Zha, 2015; Villari, Abdulla, Sebastian, & Ramaprasath, 2020).

For information retrieval applications, MAB evaluates the predicted relevance of documents and selects them while taking into account the uncertainty in the assessments by adaptively selecting the most relevant depending on the feedback that is obtained (Losada, Parapar, & Barreiro, 2017).

In energy management, by dynamically choosing the optimal action based on feedback received from the system, the MAB-based approach reduces energy consumption while improving system performance, allowing for efficient load balancing, demand response, and energy trading in smart grids and microgrids (Tran-Thanh, Rogers, & Jennings, 2012).

Within the aviation area, it can be used to make decisions about whether to climb, maintain altitude, or descend based on feedback received from sensors that measure atmospheric conditions (Bird & Langelaan, 2020)

MAB is also a useful approach in e-commerce for optimizing product recommendations and advertising. This involves selecting the most relevant products or ads to display to each user based on their behavior and preferences, with the goal of maximizing the likelihood of a purchase. MAB algorithms work by exploring different options, such as products or ads, and then exploiting the ones that are most effective in generating sales. By continually learning from user feedback, MAB can improve the relevance and effectiveness of recommendations and advertising over time (Chakrabarti, Kumar, Radlinski, & Upfal, 2008; Liu & Li, 2021).

For social media influence the goal is to identify the most effective individuals to promote a product or idea. MAB algorithms can be used to select the most influential users to target, based on their past behavior and the potential impact they can have on the target audience. The approach involves exploring different users to promote the product or idea and exploiting the ones that are most effective in generating the desired outcome (Wen, Kveton, Valko, & Vaswani, 2017).

By reviewing the previous studies, it is obvious the MAB can be used widely to balance the exploration vs exploitation dilemma. On this basis, this technique could be promisingly used to assist scientific user facilities to know whether additional experiments are required. Conducting

experiments in scientific user facilities often cost over a million dollars. While more experiments provide more data, they also come at a high financial burden which is substantial for its backers.

Another factor that obstructs knowledge creation is the impotence on the timely sharing of data. Timely sharing of data requires coordination among all involved entities in the related platform. In the business model, the survival of companies is contingent on the success of supply chain management. Data in scientific user facilities follow the same flow as physical products from supplier to consumer to reach exploitation. By relying on the concept of SC management, Spanaki et al. (2018) designed the DSC to guarantee timely and sufficient data sharing among its users. The optimal planning in supply chains (SC), regardless of what is going to be managed, could be determined when all information about customers' demand is known. However, when it comes to stochastic environments (Porteus, 1990), planning confronts more difficulties. The customers of scientific user facilities come from the public and private sectors and have stochastic demand. In a stochastic demand environment, the demand for a product or item is not known with certainty, and it may vary from one period to another. This uncertainty in demand can lead to overstocking or stockouts if inventory levels are not properly managed, resulting in some undesirable outcomes (Chiu & Choi, 2016). Overstocking and understocking in the setting of scientific user facilities are translated into how many more or fewer experiments have been done, respectively. The former imposes a remarkable financial burden while the latter limits the pace of scientific discovery. In the literature on supply chains, one promising strategy for confronting the complexity of the stochastic nature of customer demand is formulating the problem through the idea of the Multi-Period Newsvendor problem. The multi-period newsvendor problem is a classical stochastic inventory control problem with a long history in literature. The problem was first introduced by (Arrow, Harris, & Marschak, 1951) and works by considering historical demand data and forecasting future demand. In this direction, the retailer can make informed decisions on how much inventory to order in each period to meet customer demand while minimizing costs (Matsuyama, 2006). Via continuously monitoring and adjusting inventory levels based on demand patterns, the retailer can avoid both overstocking and understocking situations, leading to improved customer service levels (Qin, Wang, Vakharia, Chen, & Seref, 2011).

The literature on the supply chain around the Multi-Period Newsvendor problem is remarkably rich since the stochastic nature is an inseparable part of our world and inventory management is involved in many businesses (de Castro Moraes & Yuan, 2021). The application of Multi-Period Newsvendor is mainly surrounded by supply chain problems where the demand (Xiao & Wang,

2023) or price (Fathizadeh, Savinien, & Rekik, 2022) or both (Canyakmaz, Özekici, & Karaesmen, 2022; Petruzzi & Dada, 1999) are unknown.

The literature on the use of Multi-Period Newsvendor in SC around the presence of data revealed the role of data is serving in the way to understand the behavior of unknown demand (Ban & Rudin, 2019; Huber, Müller, Fleischmann, & Stuckenschmidt, 2019; Lin, Chen, Li, & Shen, 2022). Although extensive research has been carried out on the application of data to just disclose the distribution of demand in a better way in the Multi-Period Newsvendor problem in various areas such as manufacturing (Huber, Müller, Fleischmann, & Stuckenschmidt, 2019), retail (Yang, Hu, & Zhou, 2021), e-commerce (Qiao, Yang, & Sun, 2022) healthcare (Görgülü & Sarhangian, 2022) no single study exists which employ Multi-Period Newsvendor idea on managing the inventory of data as the deliverable product.

Given the complexity and multi-disciplined nature of the SDSC problem, consulting the multi-period newsvendor (Matsuyama, 2006) model is advantageous because it highlights the ability to share information and adjust demand more frequently, increasing the number of ordering periods during a single selling season. This model produces higher rewards for the retailer and overall supply (Wei & Chen, 2021)

Several facilities within the U.S. Department of Energy and NASA are working toward the real-time steering of experiments (Noack & Zwart, 2019; Bicer, et al., 2020; Levental, et al., 2020; Elias, Chard, Libera, Foster, & Chaudhuri, 2020). While experiments generate a great deal of data, the actual number of scientific breakthroughs has not kept pace with the data's promise. By promptly disseminating data, customers can capitalize on it and advance scientific discoveries. In the real world, the exact demand distribution for data during a test campaign at a scientific user facility is unknown. Requiring a high amount of financial support, the customer always struggles with the conundrum of exploration vs exploitation, and they must know to which extent conducting more experiments will result in maximum expected reward. Reward in this realm could be translated into knowledge creation.

Based on our knowledge and research, this is the first study on designing a SC framework with the aim of accelerating scientific discovery. To this end, through hybridizing the multi-period newsvendor model and multi-armed bandit framework, we proposed a hybrid approach, the *Multi-Armed Newsvendor Model* to take the benefit of both in the regard of accelerating knowledge creation. The multi-period newsvendor model emphasizes the importance of information sharing, while the multi-armed bandit framework instructs retailers on how to best deal with the exploration vs exploitation dilemma. Exploration is expensive while testing at a scientific user facility. The

customer has a finite set of budget and time allotted to test their hypotheses. The multi-armed newsvendor model guides the customer in exploring in a cost-effective measure while gathering enough information to prove or disprove their hypotheses. It is imperative experimental data is shared among the cross-discipline team while actively testing to heuristically make informed decisions.

Methodology and Solution Approach

In this research project, our primary objective is to tackle the challenges customers face while testing at scientific user facilities, particularly their inability to fully utilize the vast amounts of data generated, resulting in a slower pace of scientific discoveries. To achieve this goal, we propose a new framework inspired by the data supply chain model, which aims to improve the efficiency of data management, processing, and analysis in scientific user facilities. This framework assists scientists to extract valuable insights from the spectacular amounts of data generated by these facilities, leading to more discoveries and breakthroughs in various fields of science. Our study aims to provide a comprehensive solution to the existing challenges and promote the use of scientific user facilities to their full potential.

The U.S. Department of Energy and NASA oversee several scientific facilities that are actively striving to implement real-time experiment steering, where experiments can be monitored and adjusted in real-time (Noack & Zwart, 2019). It would be beneficial to develop an abstract model that outlines the advantages of transitioning from current operations to next-generation, data-driven operations in premier scientific user facilities and integrating digital transformation into standard operations. Leveraging concepts from the supply chain management field, hybridizing two mathematical models, namely the multi-armed bandit and multi-period newsvendor models, resulting in the *Multi-Armed Newsvendor Model* shows promise in this regard. By applying these models, we can make better-informed decisions when dealing with the vast amounts of data generated by advanced experimental facilities, leading to improved efficiency and better outcomes.

The next three subsection are as follows. Subsection one elaborates on the multi-arm newsvendor model and proposes a solution strategy for accelerating scientific discovery. Subsection one explains the theoretical model for knowledge creation to display how the multi-armed newsvendor model works. Lastly, Section 3.3 comprehensively explains how designing scientific data supply chains could speed up the pace of scientific discovery.

Multi-Armed Newsvendor Model and Solution Strategy

The explore-then-exploit strategy is a common approach to decision-making that involves first exploring all possible options and then exploiting the best option once it has been identified.

This strategy can be effective in situations where there is a lot of uncertainty about the best course of action, but it can also be costly, as it requires exploring all possible options before making a decision.

In the context of large scientific user facilities, exploration can be extremely expensive, and therefore, it is important to develop a new framework that can guide customers when testing to understand complex phenomena at large scientific user facilities. This framework should consider the cost of exploration and should heuristically make decisions based on the data as it is shared.

To do so, we proposed a *multi-armed newsvendor model*, which has been formulated through hybridizing the multi-armed bandit and multi-period newsvendor models. The former seeks to establish balance between exploration vs exploitation; however, the latter can be used to make exploration cheaper and more efficient by allowing for shorter time intervals and smaller amounts of inventory. The synergy that comes from hybridization empowers the proposed model by significantly augmenting the benefit of their mutual impacts in the regard of accelerating scientific discovery.

Our proposed solution strategy for successfully employing the multi-armed newsvendor model would be twofold: (I) The user wants diversity during the exploration phase (Benureau & Oudeyer, 2016; Durand, et al., 2018) especially if the reward distribution is truly unknown; (II) During that exploration phase, the user would like to gain as much information about reward distribution as possible. This can be achieved by reducing the dimensionality of the data to gain more information at a lower cost and by demonstrating superior results (Shen, Wang, Jiang, & Zha, 2015).

Figure 3 displays the framework for the proposed hybrid model in the context of scientific user facilities. At scientific user facilities, customers typically have a set amount of time and money allocated for their experiments. To make the most of these resources, customers often come with multiple hypotheses to test, prioritized by importance. It's important for customers to use all of their allotted time and money during the experiment, so they must have enough hypotheses to test in order to avoid wasting any resources. The generated data will then be used to either prove or disprove each hypothesis. The framework has two consecutive phases: exploration and exploitation. The customer will begin with the highest priority hypothesis to test. Experiment i will be performed, generating data that is then shared with the knowledge bins. Within each of the

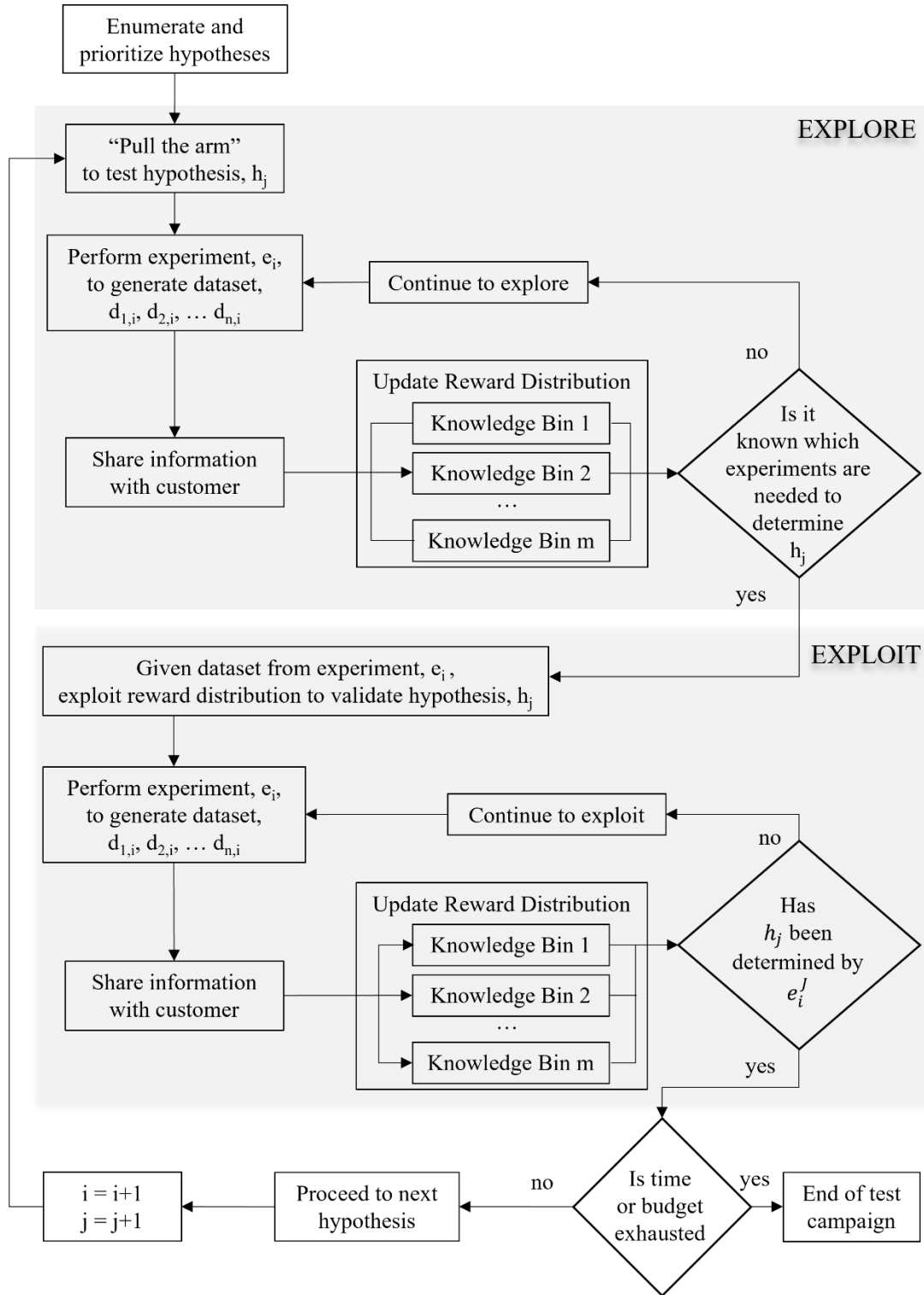


Figure 3: Multi-Armed Newsvendor Model for Scientific User Facility

knowledge bins, scientific computational models are used to convert data / information to knowledge, and the models are refined based on the data. Each knowledge bin represents a unique discipline that holds some responsibility towards the final product. The reward distribution represents a given discipline's model and whether it is known if they can predict the likelihood of an outcome for a given test condition. After the reward distribution is updated, it is assessed whether it is known which experiments are required to determine the current hypothesis. If the experiments are not known to determine hypothesis j during the exploration phase, the customer must continue to explore. If the required experiments are known, the customer proceeds to the exploitation phase.

In this phase, a new experiment is defined, a new data set is generated, information is shared, knowledge bins are updated until hypothesis j has been determined. Once hypothesis j has been determined, the remaining time and budget is assessed. If resources have been exhausted, then the test campaign is over. If resources are still available, the customer proceeds to the next hypothesis and begins the exploration phases again.

It is important to understand that although the rewarded distribution is unknown at the beginning of the experiment, the data collected during the test campaign is subsequently made available for the remaining duration of the experiment. Additionally, the number of hypotheses and knowledge bins does not always correspond on a one-to-one basis. The quantity and rate of information shared among researchers will vary based on the type of experiment, the facility's resources, and the specific hypothesis being tested. The main objective of this framework is to take a comprehensive approach in determining the successful establishment of relationships between parameters, thereby optimizing the acquisition of knowledge.

The multi-armed newsvendor model would be helpful in many applications, especially for unique, expensive experimental campaigns studying the frontiers of science where the time in the facility is fixed, large amounts of data are collected, and informed decisions must be made quickly to maximize the reward of knowledge creation.

The Department of Energy's Accelerating Scientific Knowledge Discovery Report (Department of Energy, 2013) highlights the importance of processing and sharing data as quickly as possible. Mapping these challenges to a supply chain, it becomes clear that the real value of quickly processed data is only realized when it is shared as quickly as it is processed. Hence, the use of supply chain theory has the potential to guarantee the achievement of scientific user facilities in using the full potential of data.

Theoretical Model for Knowledge Creation

Scientific user facilities aim to provide high-quality research facilities and systems that enable customers to achieve the highest reward distribution. Typically, customers have a set of hypotheses that they need to test and plan a sequence of experiments before using the facility, with the goal of maximizing their ability to validate or invalidate their hypotheses. However, it may not always be clear what experiments need to be performed at the start of the experiment. Due to the high demand for scientific user facilities and the complexity of the scientific challenges they tackle, it's often not possible to address all hypotheses with a single test campaign. To address this, we propose a framework that minimizes the cost, measured by the number of unique test campaigns, required to generate sufficient data to address each hypothesis. A test campaign is considered a unique entry into the scientific user facility, and multiple test articles can be assessed within a single campaign. However, customers must leave the facility once they have used up their allotted time or budget, and a new test campaign is required for further testing.

Let H represent the set of hypotheses that must be addressed and $h_j \in H$ represent a given hypothesis. Let E represent the set of experiment types that can be performed, where $e_i \in E$ represents a specific type of experiment. We say that hypothesis h_j can be determined (validated or invalidated) if the appropriate number of experiments is performed. That is, for every experiment e_i , there is a value e_i^J such that h_j can be determined if at least e_i^J tests have been performed for each e_i . In this context, we assume that the specific values of e_i^J are unknown to the user at the start of the experiment. However, as experiment e_i is performed, the accuracy of the user's prediction of e_i^J increases, so when there is sufficient data to determine h_j , the user is aware of this. The proposed formulation has been inspired by *the multi-armed news vendor model*. Table 2.1 provides the indices, parameters, and variable with their descriptions for the proposed formulation. Table 2.2 outlines the formulation of the problem.

Table 1: Notation and Description

Indices	
i	Experiment
j	Hypothesis
t	Test campaign,
Parameters	
$a_{i,j}$	How much data experiment i provides for hypothesis j
h_j	Amount of information needed to satisfy hypothesis j
T	Upper bound on number of test campaigns
d_t	Amount of time allotted for test campaign t
Variables	
f_t	1 if test campaign t is performed, 0 otherwise
$e_{i,t}$	How many or how long experiment i is performed in time t

Table 2: Problem Formulation

Equation	Equation Number
$\text{Min } Z = \sum_{t=1}^T f_t$	(1)
$\sum_i a_{ij} e_{ij} \geq h_j \quad \forall j$	(2)
$\sum_i e_{i,t} \leq f_t d_t \quad \forall t$	(3)
$e_{it} \geq 0, f_t \in \{0,1\} \quad \forall i, t$	(4)

The goal of this model as shown in equation (1) is to minimize the number of test campaigns which indirectly results in minimizing the cost. Constraint (2) ensures that enough information is generated to satisfy each hypothesis. Constraint (3) ensures that all experiments can be feasibly performed in each test campaign. The last one bounds the variable. In our formulation, the h_j terms are unknown; however, the more data is collected towards advancing hypothesis j , the more information we know about h_j . Moreover, we always know when we have enough information to determine hypothesis j . We treat this as a multi-stage problem, meaning that experiments in test campaign t can be determined after the data from experiment $t - 1$ is known.

As a means of gaining insight into the model's operation from a cost-cutting perspective while simultaneously maximizing the number of tested hypotheses, we run a heuristics simulation.

A reference time of 160 hours is used. It is assumed there are three hypotheses to satisfy. For each hypothesis, an initial guess for number of experiments to satisfy each hypothesis j is determined from a uniform distribution, $h_j = \text{Uniform}(0, 100)$. Each hypothesis j has unknown demand, d_j , with a normal distribution with mean, μ_j , and standard deviation, σ_j initially equal to h_j , $d_j = \text{Normal distribution}(\mu_j, \sigma_j)$. An experiment is ran and data are assigned proportionally among the hypotheses. Data are assigned to each hypothesis, h_j . The information is shared, and for each hypothesis, the unsatisfied number of experiments is calculated. The demand is updated, and a new h_j is calculated by subtracting the previous demand, d_j , from the assigned data: $h_j = d_j - r_j$. It is evaluated if all hypotheses been satisfied. If not all of the hypotheses have been satisfied, then a new experiment is ran. If all hypotheses have been satisfied, then the simulation is finished. For each, several time intervals are ran, from five hour time windows to 160 hour time windows, in five hour increments. For each simulation, there are 35 results stating the number of hours required to satisfy all hypotheses. With respect to the reference point, 160 hours, a percentage of improvement is calculated. Figure 4 depicts the flow of the simulation's logic. Figure 5 reveals the results from running 1000 simulations.

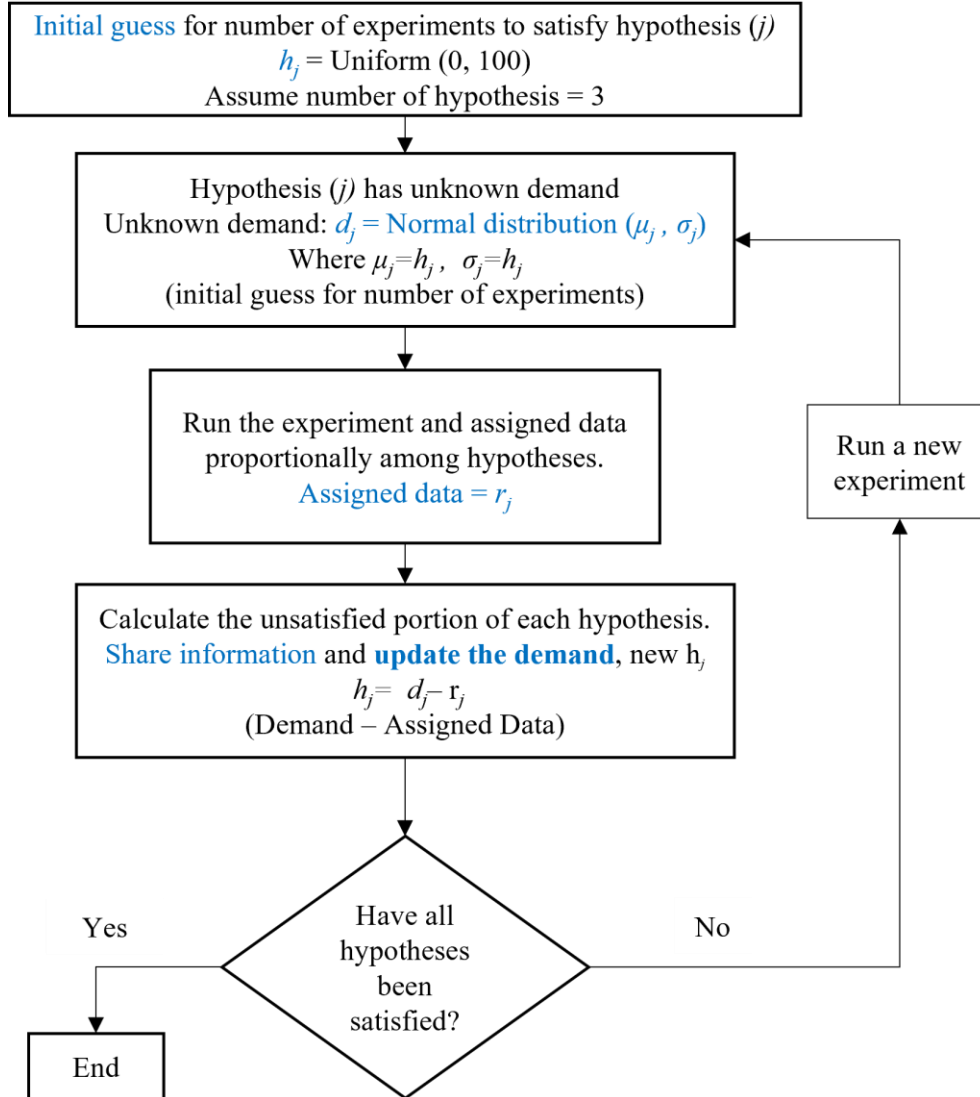


Figure 4: Simulation Logic Flowchart

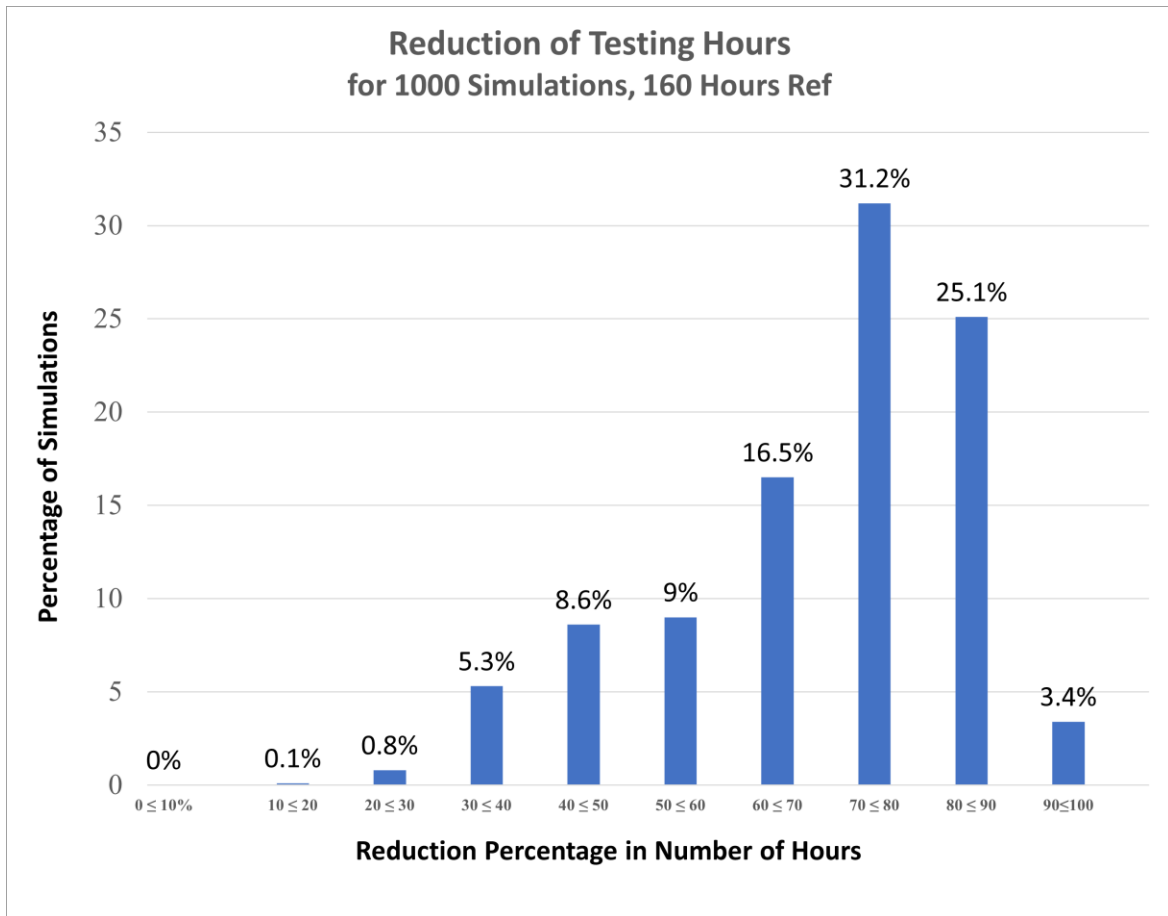


Figure 5: Multi-Armed Newsvendor Simulation Results

For 1000 simulations, the average improvement observed was 70.2%. This means on average, test campaigns were reduced by 70%, from 160 hours to 48 hours to satisfy all the hypotheses. The minimum improvement observed was 18.8%, which can be interpreted as reducing the test from 160 hours to 130 hours. The maximum improvement observed was 95.3%, which can be interpreted as reducing the test from 160 hours to eight hours.

Scientific Data Supply Chain

This study is focused on finding a solution to assist scientific user facilities fully utilize the potential of their generated dataset. The two previous subsections have elaborated on our proposed hybrid approach and manifested how *the multi-armed newsvendor model* could cost-effectively enhance the scientific discoveries in an uncertain environment. Realizing the benefit of the proposed hybrid model requires to be implemented on the efficient infrastructure and supply chain theory could be helpful to this end. Roth and Singhal (2022) declared that to use the fullest potential of scientific data, it is required to share as widely and timely as possible. Due to the multidisciplinary nature of data, if it became available for wide audiences, scientific discovery could occur simultaneously in multiple disciplines. This is what exactly the current study is looking for, accelerated scientific discovery.

To better demonstrate the weakness of the inefficient infrastructure of scientific user facilities which has limited the pace of scientific discovery and shed a light on promising solution, we have taken a look at the recent history of NASA's design and development of the Space Launch System while testing at the agency's scientific user facilities (i.e., wind tunnels) for conducting a series of advanced experiment assessing the performance of the rocket against high spatiotemporal unsteady, chaotic flow. We initially discuss how the inefficient infrastructure presented obstacles to NASA while testing at the scientific user facilities. Then we outline a resolution for attenuating the scientific user facility's system weakness.

Traditional scientific data supply chain - NASA began development of the Space Launch System (SLS) rocket in 2011. The SLS launch vehicle is the world's largest rocket ever developed and flown and will launch humans further into space as part of NASA's Artemis Program. In 2012, NASA began testing the SLS's performance against unsteady, chaotic flow. Dynamic aerodynamic loads dictate a launch vehicle's structural design (Piatak, Sekula, Rausch, Florance, & Ivanko, 2015). Understanding, predicting, and managing high spatiotemporal, unsteady flow is the most pressing challenge facing the field of aero-sciences and virtually affects the entire discipline. To

understand chaotic, unsteady flow, it is critical for the customer to measure fluctuating surface pressures experimentally and computationally, but the ability to do either is challenging.

After the initial wind tunnel tests in 2012, engineers produced results that estimated the aerodynamic load, introduced by the chaotic, unsteady environment the SLS rocket would fly through, would not meet program requirements. Understandably, this caused significant concern that an extensive redesign would be required. A special task team was assembled to address why these results were found and identify options to mitigate high aerodynamics load estimates. This included examining the models used to estimate the aerodynamic loads from the original test and examining new hardware to mitigate the aerodynamic loads (Sekula, Piatak, Rausch, Florance, & Ramey, 2015). Fortunately, during a previously scheduled wind tunnel test eight months later in 2013, these mitigation options would be assessed, and then again six months later, in 2014, when the SLS Program requested high priority access to the scientific user facility. During these tests, additional instrumentation was added to increase the sensor's spatial resolution measuring the unsteady surface pressure fluctuations, and new numerical models would be produced based on the higher resolution data. Both solutions were expensive options for NASA. For this effort alone, the approximate cost was \$2 Million for additional testing and \$1.5 Million for the special task team. Neither of these includes any cost associated with schedule delays, additional workload imposed on NASA and its commercial partners, or increased travel requirements.

It should be mentioned that the intent of this case study is retrospective and is not a critique to the NASA SLS program or NASA's scientific user facilities. The SLS Program was experiencing the same trouble the aero-sciences field had experienced for decades (Schuster & D'Agostino, 2017). As a result, the NASA Engineering and Safety Center (NESC), NASA's independent engineering and safety organization, executed a wind tunnel test to assess the usability of a new high-resolution optical technique to globally measure fluctuating surface pressures, called unsteady pressure-sensitive paint in 2015 (Schuster, et al., 2016). It is not until the results of this research test were processed did the SLS Program know how to correctly estimate the unsteady loads produced by the chaotic, unsteady fluid environment, approximately four years after the original test in 2012. Two things were wrong with the initial SLS wind tunnel tests; first, higher spatial resolution data were needed to understand the complex flow phenomena. Second and most important, data needed to be delivered in near real-time for customers to assess their models' efficacy.

Through empirical observation in related NASA scientific user facilities, we reached Figure 6. As highlighted by Bear et al. (2021), the scientific user facility is an embedded system, shown by

the bold line in Figure 6, and is severely spatially limited since it hosts many functions. Within the scientific user facility, the facility control system, facility crew (10–20 people per shift), instrumentation suites (Instr. Suite), local computational resources, and a limited number of customers are all hosted inside the physical walls of the experimental facility. The facility control system is the main system that automates and controls the operation of the facility. The facility crew will have a broad range of the required experts: test engineers, electricians, mechanics, instrumentation technicians, and owners of various instrumentation suites.

Many of these facilities have bi-weekly turnaround of customers, each with a new test article and test objectives (Thayer, et al., 2019; Roozeboom & Baerny, 2017). Approximately 10 – 12 tests are conducted per year at many of NASA’s facilities. Several weeks after the test is complete, data is loaded onto an encrypted hard drive and mailed via a traditional mail courier. Once the customer receives the final processed data, the data are distributed among the customer's team. The larger team of the principal investigator (PI) and domain experts (D1, D2, ... Dn) and individual contributors will work together to arrive at a final assessment of their test campaign. Depending on the stakeholders and the complexity of the test, this can take several months. If negative results are found and the test campaign requires modification and retesting, they must make a new request to the scientific user facility. Due to high demand, the customer may be required to wait several months or even up to a year to retest. While this is not common for every test, a customer designing a new state-of-the-art vehicle or material will most certainly repeat some portion of a test, increasing the instrumentation and editing of the test matrix, since they are on the frontiers of science where models and simulations are not well established.

Viewing Figure 6 through the lens of supply chain management, there are improvements that would improve the efficiency of the traditional scientific data supply chain. The length of a supply chain can be expressed in terms of its lead-times and number of stages. Longer supply chains are more vulnerable to disruption and longer response times to recover (Graves, Tomlin, & Willems, 2022). The length of the data supply chain in Figure 6, is an indicator of the vulnerability posed to the customer and the scientific user facility. Since data is the artifact moving through the supply chain, more attention should be dedicated towards decreasing the lead time of data being delivered.

By examining how the supply chain is structured, we have revealed areas where the scientific user facility can most effectively and efficiently innovate (Carter, Rogers, & Choi, 2015). Figure 7 shows the visible horizons within the manufacturer (scientific user facility) and the customer.

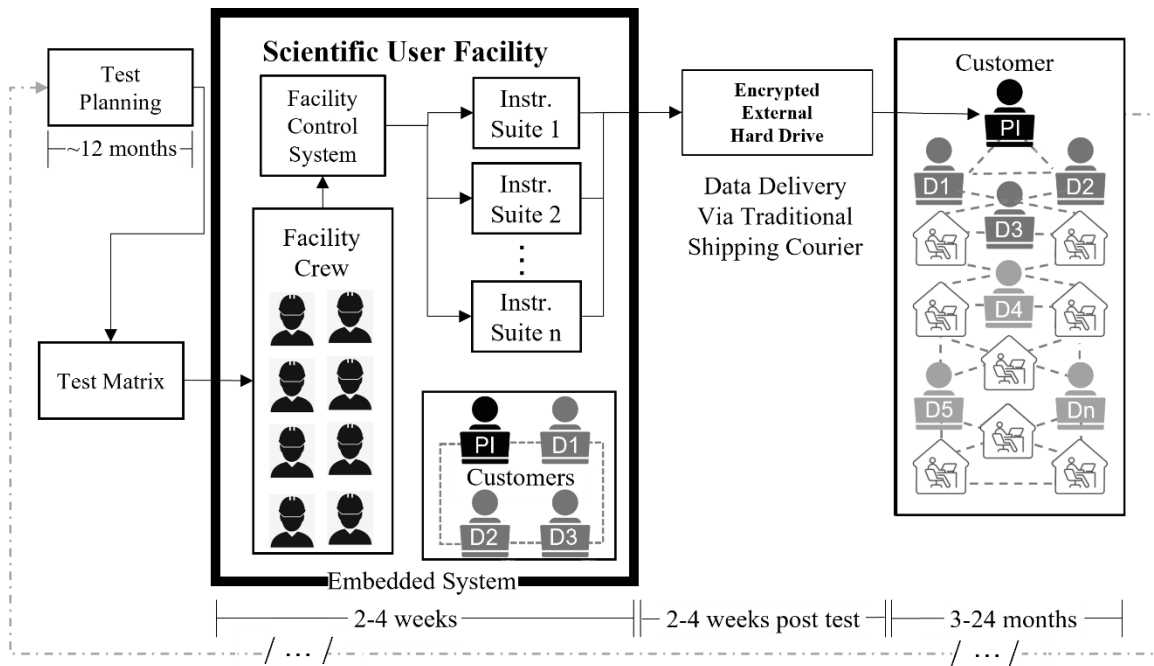


Figure 6: Traditional Scientific Data Supply Chain

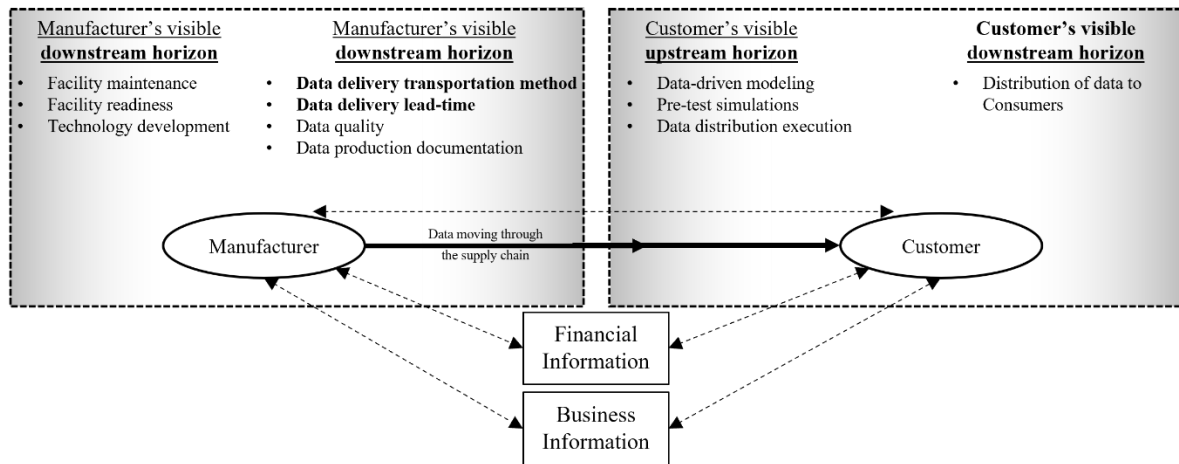


Figure 7: Manufacturer's and Customer's Viable Upstream and Downstream Horizon

These horizons should be set within the concurrent limits of an organization's control. The data delivery method and lead-time are within the visible downstream horizon for the manufacturer. Therefore, the scientific user facility has exclusive control over the data delivery transportation method and data delivery lead-time. The customer must simply live with what is outside of their control. Figure 7 reveals the scientific user facility can effectively innovate with how and when the data is delivered downstream to the customer.

An influential aspect that should not be overlooked, as non-profit organizations the scientific user facilities are not responsible for making sure each customer gets the most out of their investment as far as science goes. But the facility is responsible for creating an environment in which each customer can do so.

Transformational scientific data supply chain – In 2019, NASA conducted a first-of-its-kind demonstration called Project, Red Rover (NASA, 2019). Red Rover successfully connected the data stream from an experimental facility, Unitary Plan Wind Tunnel (NASA, 2022), to a computational facility, NASA Advanced Supercomputer (NAS) Pleiades (NASA, 2022), the 90th most powerful computer in the world (TOP500, 2022). The UPWT and NAS Pleiades are co-located at the NASA Ames Research Center in Moffett Field, CA. These two facilities are managed under different mission directorates within NASA and share very little overlap due to their funding and organizational structures.

However, the field of fluid dynamics is one of the first big data fields (Mendez, Ianiro, Noack, & Brunton, 2022). High-performance computing began after World War I, with the need for more computational resources required for the design of weapons and aircraft (Senning, 2020). With the increasing quantity and sophistication of sensors and optical technologies, UPWT has become a big data generator with insufficient computational resources within its current capabilities. Due to security restrictions, many scientific user facilities do not have a network connection to the outside world. Inside the facility, the computers and sensors are connected on a secure, isolated network. However, hardware and software now exist to allow secure communication from one network to another. During the Project: Red Rover demonstration, 150 terabytes of data, collected over four days, were transferred between facilities, processed on the supercomputer, and distributed to customers in near-real-time during an experimental test campaign. This demonstration would mark the entry into a new phase of experimental testing at scientific user facilities at NASA using a transformed scientific data supply chain, as depicted in Figure 8, empowering customers and facility managers to make data-informed decisions.

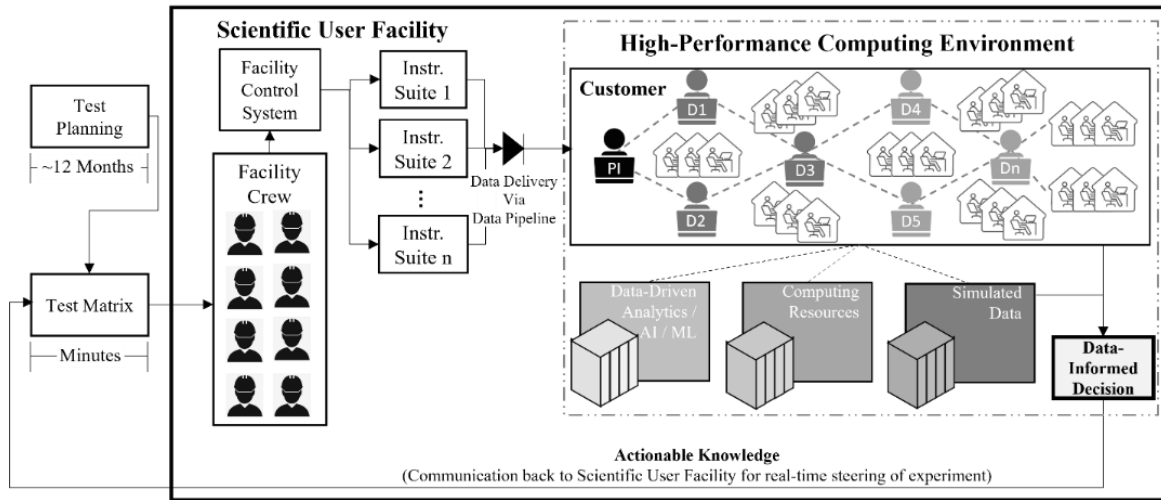


Figure 8: Transformational Scientific Data Supply Chain

For this digitally transformed scientific data supply chain, test planning would still require about a year, but more emphasis is placed before the test is executed on the algorithms and models that will be employed to enable the real-time steering of an experiment by the customer. A key difference between Figure 6 and Figure 8 is that the physical constraints of the scientific user facility have been removed from the transformational data supply chain. The scientific user facility has expanded to include access to a high-performance computing environment (or similar) to transform this embedded system into a cyber-physical system.

The customer has access to and is part of this high-performance computing environment, which allows access to HPC enterprise-level tools (like data-driven analytics), complementary datasets (like computationally simulated datasets), and unequivocal computational resources. The number of customers is not limited by the physical space in a room. This network of customers coupled with HPC tools enables the discovery of actionable knowledge that enables informed decision making and the real-time steering of an experiment. The test matrix is updated, and the facility control system executes the desires of the customer. The success of the test campaign and the goal of maximizing knowledge creation during the test campaign are key elements of the framework.

By mapping these big data challenges to a supply chain, as shown in Figure 6, we exploit two valuable tools from supply chain management to solve the challenges limiting knowledge creation: transportation method and information sharing. The value created by a firm is accomplished by delivering a specific product and / or by the value is delivered (Bals & Tate, 2017).

One of the most significant factors in determining the transportation method of a product is the value-to-weight ratio. Products with a high value-to-weight ratio should be moved as quickly as possible. Consulting the hybrid multi-armed newsvendor model, the newsvendor model provides motivation to move the product (scientific data) from each specific experiment, e_i , as quickly as desired, to allow the user to know when a given hypothesis, h_j , is determined. Given the extensive planning (~1 year) and high cost for a single experimental campaign (~\$2,000,000 or more), data should be transported as quickly as possible to assess if h_j has been determined by e_i . Electronic delivery through a data pipeline is the only transportation method that can meet value-to-weight requirements. If the data are processed and shared through a platform such as an HPC facility, the number of customers with direct access to the data without delay will increase. However, simply moving the data faster from one location to another only solves transportation challenges. Moving the data to a location with accessibility to more individuals in real time with enterprise-level tools

for sharing and data analysis exposes the value of information sharing in a supply chain. The benefits of information sharing give us the motivation to improve transportation methods.

Information sharing demonstrates the broader, holistic perspective that modern supply chain management research has highlighted (Wieland & Durach, 2021) and enables the supply chain to improve communicate and anticipate variations, permits greater efficiency and planning by smoothing product manufacture and decreasing inventories (Chatfield, Kim, Harrison, & Hayya, 2009). Due to a lack of information, each echelon highly influences the supply chain by ordering too much or too little, which leads to the bullwhip effect, a distortion and amplification of demand information within a supply chain generally causing an increase in production, inventory, capacity planning, and transportation costs. The most effective way to reduce the bullwhip effect is providing each stage with more accurate and timely information, which lessens customer demand uncertainty and lead time along the supply chain (Chen, Drezner, Ryan, & Simchi-Levi, 2000) reducing vulnerability (Graves, Tomlin, & Willems, 2022). Research conducted by Li et al. (2016) and Islam et al. (2022) concludes that variability across the supply chain is lower when demand information is shared.

From the multi-armed newsvendor model, the user must learn the reward distribution of each knowledge bin. The user can only determine if the prediction, e_i^J , is sufficient to determine if h_j is validated or invalidated through information sharing across the data supply chain during the test campaign. The user will continue to run experiments that may or may not reach the value of e_i^J . If information is not shared during the post-test analysis, the user will learn that h_j has not been sufficiently tested by experiments, E , and must retest with additional experiments. Information sharing can avoid this negative result.

Given that the primary goal of these scientific user facilities is to generate new knowledge, information sharing is essential for accelerating the pace of scientific discovery. In a knowledge-service setting, “productivity is measured as the amount of information exchanged per unit time” (Jónasson, Ramdas, & Sungu, 2022). Real-time information can reduce uncertainty and increase the production of higher profits (knowledge creation). Transforming the scientific data supply chain to include a data pipeline shares experimental data with more people using enterprise-level tools like HPC. Hence, more individuals have access to experimental data and powerful tools to aid in data analysis and decision making. This enables a user to have actionable knowledge that enables real-time, data-informed decisions, ultimately accelerating the pace of knowledge creation.

Through information sharing, a better product is produced because the reward distribution is constantly assessed and updated.

Discussion

Exploring the frontiers of science is challenging and requires a large amount of data to be collected for cross-discipline teams to assess. However, scientific user facilities are embedded systems and do not easily allow for customers to share these valuable data sets.

By retrospectively analyzing previous tests conducted by NASA's SLS Program at NASA scientific user facilities, we are able to confirm that the lack of information sharing while testing greatly hampered the teams' ability to prove or disprove their hypotheses and make data-informed decisions, which in this case caused the program to add additional tests to their schedule, costing the program millions of dollars and years in schedule.

We know from existing literature, the MAB model is an intuitive, cost-effective framework for addressing the exploration versus exploitation dilemma, which all customers face when exploring the scientific frontier while testing at a scientific user facility. We also know from research studies on the multi-period newsvendor model the ability to share information and adjust demand more frequently produces higher rewards for the overall supply chain. Therefore, by combining these two models with the goal of accelerating scientific discovery, a customer is able to cost-effectively iterate faster through hypotheses by sharing information while testing at a scientific user facility.

Currently this research is limited to studying previous cases where a digitally transformed scientific data supply chain would have benefitted the customer early in the design cycle. New opportunities are forthcoming, but digitally transforming a scientific user facility takes time. It also requires preparation and research on the customer's side to realize the benefits of leveraging HPC to update models for near-real-time steering of their test campaign.

The potential of the multi-armed newsvendor model and digital transformation of the scientific data supply chain to make a significant impact is vast and varied. We delve into this impact where we explore the benefits in the realms of scientific research, facility and operations management, and engineering management, respectively. By examining each of these areas in detail, we can uncover the full spectrum of advantages this model has to offer.

Potential Benefit for Scientific Research

In the world of scientific research, the exploration versus exploitation dilemma can be a difficult one to navigate. It's important to explore as much as possible but doing so can be costly. The innovative, hybrid multi-armed newsvendor model is an impactful framework that allows

researchers to explore economically, which is especially useful when working on the frontiers of science.

The Department of Energy's Advancing Scientific Knowledge Discovery Working Group report highlights the importance of this framework in scientific research. As hypotheses become more complex and cross-disciplinary, subject matter experts need to be closely connected during the scientific evaluation. The multi-armed newsvendor model addresses the need to make data available quickly and to a broad, cross-disciplinary audience, regardless of their location.

This model is essential for understanding the complex evolution of a phenomenon through a series of decisions. Exploration is necessary to understand a problem, and exploitation is key to exploiting incremental peaks and gradients based on the knowledge created during the exploration phase. In short, the hybrid multi-armed newsvendor model is a powerful tool for scientific research, allowing researchers to explore and exploit as efficiently and effectively as possible.

In the field of aerosciences, the vulnerability of the scientific data supply chain was exposed during the development of NASA's Space Launch System. The approach to add instrumentation and retest is not preferred or sustainable, as it constitutes an extremely expensive method to exploration. Later in Schuster and D'Agostino (2017) report, they discuss innovative technologies that began development because of the decades of testing that continued to produce familiar, yet unfavorable, results. Several of these technologies are optical and follow the trend highlighted previously that many fields are moving to optical-based technologies. The broad field of scientific research uses optical-based technologies to study and continuously improve their understanding of complex evolution over space and time. Even with additional technologies, alone, scientific discovery would have required months for processing and delivery because sufficient computational tools to store, process, and share data in an efficient manner are an essential part of the data supply chain. Optical technologies quickly increase data volume and velocity, amplifying the requirements for a digitally transformed process to effectively reap the value of these data and accelerate scientific discovery. By relying on Figure 6, in each case, if the team executing the tests had access to data in real time and high-performance computing resources, better decisions could have been made. Ultimately, the retesting of 2014 could have been avoided, saving at least \$4 million for this single example.

With respect to medical research, our proposed model bears an intriguing advantage worthy of attention. Exploring the world of medical research can be a daunting and perilous journey. With the costs of clinical trials on both humans and animals skyrocketing, it's crucial to find ways to navigate the terrain without compromising the quality or quantity of information gathered. Kuleshov and

Precup's (2000) research underscores the vital role of multi-armed bandit algorithms, which were initially conceived by (Gittins & Jones, 1979) as a solution to the daunting challenge of clinical trials. Meanwhile, (Coutant, et al., 2019) explain that building systems biology models is an uphill battle due to their complexity and relatively low performance. Even a seemingly simple model like *E. coli* contains thousands of molecules and proteins that interact in intricate spatiotemporal ways. To enhance these models, hundreds or thousands of cycles of analysis are needed, a tall order given that most studies can't even complete one cycle. The solution, as the article concludes, lies in closed-loop cycles of data acquisition, processing, and sharing. To make this possible, automation and near real-time information sharing are essential for scientists to make informed decisions about which experiments to execute next. The key takeaway from the research is that effective exploration based on data is much more effective than wandering blindly, and the multi-armed newsvendor model provides a practical framework for putting these insights into action.

Potential Benefit for Facility and Operations Management

Earlier, it discussed the ongoing tension faced by both not-for-profit and for-profit operations managers. They must navigate the delicate balance between using revenue for essential maintenance and investing in new technology to attract new customers to meet the evolving needs of their customer base. Despite these challenges, managers must also consider the potential benefits of redesigning their supply chain to better serve their customers and improve facility operations, even when the payback is uncertain.

Operations management insights are crucial for the development of new models, such as the multi-armed newsvendor model. Facility and operations managers must stay attuned to their customers' needs and successes, recognizing when repeat customers are not effectively utilizing their services. If the goal of the not-for-profit operations is genuinely social merit, managers must view such inefficiencies as a red flag. By implementing technological and organizational innovations and continuously assessing the system, large system dysfunctions can be reduced.

Moreover, with the rise of technology, workflows and operations must be rethought. To accelerate scientific discovery, automating scientific workflows is essential, as stated by (Coutant, et al., 2019). The adoption of information technology can enhance automation without the same challenges associated with the production of physical goods in for-profit operations. Rather, automation would decrease the workload of constrained resources

within not-for-profit operations. However, any new technology that improves and changes standard operating procedures requires the workforce to be reskilled and trained to reap the benefits of automation for the facility and the customer.

Bals and Tate (2017) emphasized the challenges that managers face in making these decisions, but insights from operations management can provide guidance in designing new models and optimizing facility operations. Similarly, KC, Scholtes, and Terwiesch (2019) highlighted the importance of staying aware of customer needs and successes to inform these decisions. Kalkanci, Rahmani, and Toktay (2019) also contributed to the conversation with insights into the benefits and challenges of automation.

Potential Benefit for Engineering Management

“Digital transformation is more about the people than the technology”, (Hill, Le Cam, Menon, & Tedards, 2022). Today's engineering managers must be able to direct technical work, but equally important, they must continue to train themselves in how to manage the people under their supervision. The human actor in the technical organization is what makes the technical organization work and the reason for the organization's existence. Digital transformation includes the development of an organization's mindset and outlook that values valuable skills like inclusiveness, openness, and teamwork (Deloitte, 2017). The workforce has changed post-2020. Today's engineering managers should be aware and responsive to innovative technologies that allow their organizations to improve and adapt to the post-2020 workforce landscape.

In the wake of the pandemic, the workforce has undergone a remarkable metamorphosis, where colleagues are able to connect with each other effortlessly and seamlessly, no matter where they are located. Meanwhile, the scientific user facility wields the power to revolutionize the delivery of data to their clients through digital means.

For the next generation workforce, the situation is totally different. The next generation of students are entering the workforce. They experienced a once-in-a-lifetime pandemic and saw more Americans actively protesting social unrest than had demonstrated during the 1960's civil rights movement (Ruggs & Avery, 2021). This generation has been greatly influenced and is deciding how to devote their talents and genuinely devoted to making a positive impact in the world (Jónasson, Ramdas, & Sungu, 2022). Many desire to work for

organizations that value their talents and value their benefit to society. They also understand there are many organizations to work for and will not settle for poor organizational values and managers.

The next-generation workforce also strives to accomplish their ambitious career goals and continuously improve their skill sets. The well-known 10,000-hour rule goes much deeper than simply performing an action for 10,000 hours. The original research by (Ericsson, Krampe, & Tesch-Römer, 1993) titled “The Role of Deliberate Practice in the Acquisition of Expert Performance” made popular by (Gladwell, 2008), documented that the 10,000 rule requires deliberate practice, with an instructor or mentor, giving immediate informative feedback, and knowledge of the results of their performance. Therefore violin, golf, and computer programming are often examples. The trainee in any of these subjects can immediately see the results of their performance. However, for customer planning and executing a scientific test campaign with a post-test long lead time of two to four weeks, deliberate practice with immediate feedback does not exist. Digitally transforming the scientific data supply chain enables immediate feedback to the customer and in effect, better training for the next generation of scientists and engineers.

Conclusion

The current digital revolution is accelerating innovation, driving productivity, and producing more raw data than ever before, including the field of scientific research. Scientific user facilities owned and managed by the United States federal government face several challenges regarding how to systematically harness the power of big data generated. The goal of these not-for-profit scientific research facilities is to accelerate scientific discovery and increase knowledge creation by designing a system that allows customers to efficiently solve their complex scientific challenges. Through empirical observation while conducting and managing scientific research at a NASA’s scientific user facility, it was evident that the transportation method and lack of information sharing was prohibiting customers from making data-driven decisions. The data volumes were increasing, and the facility lacked the computational and network infrastructure to process and distribute key results.

Therefore, we propose a hybrid model for increasing knowledge creation at these facilities—the multi-armed newsvendor model. The power of combining the multi-armed bandit and multi-period newsvendor models, allow a multi-armed bandit model to have shorter runs during the

exploration phase and gives reason to digitally transform the data supply chain to enable real-time information sharing.

The traditional scientific data supply chain is an embedded system severely limiting the resources available while testing at a scientific user facility. With long lead times between raw data collection and final processed results, the system is vulnerable for the customer and the facility. By updating the transportation method to a data pipeline, the scientific data supply chain is digitally transformed and able to reap the benefits of the hybrid multi-armed newsvendor model.

It is the benefits of information sharing that incentivizes us to improve the transportation method, decreasing the lead time and vulnerability of the supply chain. By reviewing a series of previous tests conducted at NASA, the lack of information sharing and a digitally transformed scientific data supply chain caused the program to execute additional tests and summon additional resources. The cost to NASA was approximately \$3.5 Million for one instance.

The multi-armed newsvendor model allows researchers from the broad domain of scientific research to explore the scientific frontiers cost-effectively. A digitally transformed data supply chain also allows facility and operations managers to effectively serve their customers while improving and maintaining their bespoke facility in a data-informed way. Lastly, engineering managers can also leverage the transformed data supply chain to better train the next generation workforce while also enabling a geographically dispersed team. Future work includes identifying upcoming testing opportunities at scientific user facilities to employ the multi-armed newsvendor model.

CHAPTER THREE

GROUND TESTING 3.0: EMPOWERED TO MAKE DECISIONS

Introduction

The top, overarching technical challenge in the field of aerosciences is understanding, predicting, and managing unsteady, separate flow and practically affects the entire discipline. Unsteady, separated flow is the performance-limiting factor for many flight vehicles and dictates the structural requirements for launch vehicles and their payloads. To understand unsteady aerophysics, one must measure unsteady surface pressures, but the ability to measure these fluctuating surface pressures, experimentally or computationally, is challenging, most notably in the transonic speed regime. To address this challenge, performing a series of ground tests is one method to quantify the loads of an aerospace vehicle. The current digital revolution is accelerating innovation, driving productivity, and producing more raw data than ever before (Wozniak & Schwarz, 2019). Each industry must reflex on its goal with the increase in data available, and for scientific research, the goal is increase knowledge creation and accelerate scientific discovery. With the rise in available data, ground testing facilities are entering a new phase in which customers and facility operators are empowered to make decisions (Davenport & Kirby, 2015).

In 1949, the Unitary Plan Act was passed by the U.S. Congress and formally coordinated a plan to construct wind tunnel facilities within the U.S. Air Force, universities, and NASA's predecessor, the National Advisory Committee for Aeronautics (NACA). Due to the extensive range of requirements to respond to federal aviation needs after World War II, multiple ground test facilities with a wide range of speeds and capabilities were constructed (U.S. Department of the Interior, 2017). These facilities, along with the human capital and technology within the facility, would be essential to then-President John F. Kennedy's plans, announced to the American public in his 1962 speech at Rice University, "First, I believe that this nation should commit itself to achieving the goal, before this decade is out, of landing a man on the moon and returning him safely to the earth" (Rice University, 1962). During this time, data acquired at ground testing facilities were physically recorded by humans looking at monometer boards to record mercury levels, indicating pressure at discrete locations on the wind tunnel model (NASA, 2016). The data were graphed, and equations were derived using human computers.

Not until the late 1980s and through the 1990s would these scientific facilities see a new generation of development, Ground Testing 2.0, when many would be modernized, replacing

monometer tubes and human recorders with digital systems to improve facilities' productivity and efficiency through automation (Micol, 2001; Kmak, 2000).

Now, a third phase of ground testing, Ground Testing 3.0, has arrived with the ability to collect large amounts of data, empowering customers, managers, and engineers with the ability to make informed decisions.

Premier ground test facilities owned, maintained, and managed by the United States (U.S.) federal government carry high amounts of risk and operate solely because the U.S. federal government prioritizes the societal merit and scientific value produced by these facilities and, therefore, allocates funding for these facilities and the workforce required to run them successfully and continuously. While these facilities are owned by the U.S. federal government and managed by federal departments or agencies, such as the Department of Energy, Department of Defense or NASA, they are available to the private and public sector. The costs of the sophisticated facilities, technology, and workforce are far too great for a single private company to operate independently. Therefore, the use of these facilities is subsidized by the federal government to decrease the costs to the private and public sector to stimulate innovation, and therefore generate economic growth.

As non-profit entities of the U.S. federal government, the goal of these ground test facilities is to support their social values and maximize societal good (Tereyağoglu, Fader, & Veeraraghavan, 2017) by focusing on the “most pressing societal needs” (Bals & Tate, 2017), and for the scientific community the most pressing need is access to a state-of-the-art user facility made available to them to increase knowledge creation.

In all fields of science, the frontier is based on the evolution of a phenomenon in space and time, and future scientific discoveries will rely on a wide range of spatiotemporal scientific measurements to overcome these challenges (Stremper, Yoshii, Hammer, Bycul, & Miceli, 2021). Fortunately, these ground testing facilities also have incomparable, state-of-the-art technology that enables us to execute increasingly sophisticated experimental test campaigns by collecting data at high spatiotemporal resolutions. These measurements are often optical and capable of producing data at unprecedented volumes and velocities for the ground test facilities.

Aryal et al. (2018) claim that 50,000 gigabytes (GB) of data are collected every second, compared to 100 GB a day 25 years ago. NASA scientists, Cox and Ellsworth (1997), were the first to document this growing concern regarding challenges to manipulate large data sets and the proper computer hardware to manipulate it. This challenge is now called *big data* or *Internet of Things*.

The experimental data generated at ground test facilities fit the literature's definition of big data:

- Volume—Terabytes and petabytes are acquired in a single test campaign.

- Variety—Data are complex and multi-modal.
- Velocity—Large data volumes are acquired in days, often in multiple shifts per day.
- Veracity—Data acquired are used to design and develop state-of-the-art products.
- Value—Test campaigns are multimillion-dollar investments made by customers.

Scientific experimental data are big data by any measure. In addition to the traditional 5-Vs, Bansal, Gualandris, and Kim (2020) added that big data must also "require you to change your mind-set" (Tonidandel, King, & Cortina, 2018). In fact, the field of fluid dynamics is one of the first big data fields (Mendez, Ianiro, Noack, & Brunton, 2022). Keller et al. (2016) highlight that, if data can be integrated, queried, and analyzed, it can allow new discoveries that can impact operational efficiency, operational safety, and enable strategic decision making.

With the increased data volumes and velocities, ground test facilities have not systematically harnessed the power of big data sets (Vescovi, et al., 2020) to achieve the goal of increasing knowledge creation for their customers and increased their knowledge about the operation of their own facility. The Department of Energy's 2013 Accelerating Scientific Knowledge Discovery (ASKD) Working Group Report (Department of Energy, 2013) stated that a "new advanced discovery ecosystem" is required for the scientific research community to accelerate the pace of discovery. To achieve this, data must be processed and shared as quickly as they are acquired.

The time and effort required to harness the data grow proportionally to the data volume and variety. If these resources are not available and access to the data is delay, then the experimental data devalues with time (He, et al., 2006; Tran-Thanh, Rogers, & Jennings, 2012).

Currently, there are several contributing factors restraining our ability to accelerate scientific discover and increase knowledge creation within ground test facilities. Ultimately, ground testing facilities are embedded systems leading to an environment that is difficult to work within and inflexible (Ebert & Duarte, 2018). Secondly, the embedded environment does not allow rapid communication outside of the embedded system. A customer is limited to their own resources and not enterprise-level resources, introducing many obstacles to effectively collaborate with other team members (Lazarova-Molnar, Mohamed, & Al-Jaroodi, 2018). Third, the number of subject matter experts actively participating during a test campaign is limited, either due to physical constraints in the facility or financial constraints in the budget, which results in reducing the number of domain experts actively participating and therefore decreasing the quality of testing and innovation, increasing costs, and delaying tests (Cooke & Hilton, 2015).

These challenges, along with the large amount of data that can only be accessed within the ground test facility's embedded system, have led to a slower-than-expected pace of knowledge

creation. The data produced is beyond human interpretation, and with this data deluge, there is a severe lack of computational resources within the embedded system. Oftentimes, data is acquired but never processed, or even worse, never used. Data is essential for scientific discovery, and like any other commodity, needs careful planning and coordination across all participants to ensure quality and timely delivery.

Inspired by the idea from supply chain management, Spanaki et al. (2018) designed the data supply chain (DSC) architecture address the challenges of data exploitation. We will go one step further and define the scientific data supply chain (SDSC), which should be interpreted as a paradigm for the management, distribution, and sale of processed scientific data beginning with the acquisition of raw scientific data. Similar to a supply chain that manages material goods, the scientific data supply chain will realize efficiencies in costs when the scientific data generated during an experimental test campaign is viewed as a product to be managed. Connecting all parts of the scientific data supply chain will influence the whole system (Parkinson et al., 2020). By relying on analogical reasoning from the area of supply chain management, we give incentive to digitally transform the scientific data supply chain within ground test facilities, to harness the power of big data from ground testing facilities and empower decision makers.

Shih (2023) charged, “Doing more of the same is not enough”. The aerospace industry must reevaluate how it approaches and executes experiments in Ground Testing 3.0. To enable the growing space economy and next-generation aircraft, a wider, systematic approach must be used. Vehicles must not only be physically built using new innovative technologies, requiring less people, fewer operations, and lighter materials, but implementing new technologies during the vehicle design phase are overdue. The aerospace industry is well positioned to harness the power of big data when designing advanced aerospace vehicles by using data-driven optimization and machine learning (ML) methods (Brunton, et al., 2021) but the scientific data supply chain infrastructure must be digitally transformed to include more powerful computation hardware, improved networking, and advanced, scalable algorithms. The field of supply chain management verifies by changing the transportation method to a enable near-real-time information sharing to a broader set customer, regardless of physical location. Information sharing can reduce uncertainty and increase the production of higher profits (knowledge creation) empowering customer to make data-driven decisions.

Background

A top technical challenge in aerosciences is the prediction, measurement, simulation, and understanding of unsteady, separated flow. At low frequencies, unsteady aerodynamic forces can

couple with bending and torsional modes of the vehicle leading to undesirable responses, including loss of vehicle or even loss of crew. Quantifying unsteady aerodynamics have been a challenge for the industry since engineers began experimentally studying the phenomena (Coe, 1961). This challenge has not gone away nor becoming unimportant.

The NASA Engineering and Safety Center (NESC), an independent engineering assessment office within NASA, has invested millions of dollars over the last two decades to aid several large NASA programs in overcoming the obstacles associated with unsteady aerodynamics. In the 2017 “Future of NASA’s Aerosciences Capability” report, Schuster & D’Agostino (2017) motivate the extreme challenges and expenses due to high buffet and aeroacoustic environments caused by unsteady, separated flow for programs such as Delta II Heavy, Ares I-X, and Space Launch System (SLS). For each case, some resolution was found for the specific vehicle, and often the solution was to procure additional instrumentation, repeat the test, and / or add design margin to better understand and further refine the environments. None of these solutions were desirable, and each of these programs saw impacts to schedule and budget.

By request from the programs, the NESC would frequently assess the ability of computational fluid dynamics (CFD) tools to characterize unsteady flow, however, many of the complex, unsteady phenomena related to aircraft, rotorcraft, and spacecraft design remain difficult to predict accurately with computational simulations (Biswas, et al., 2012). Like all technologies, CFD continues to improve (Jordan & Mitchell, 2015), but currently it is too expensive, computationally, to ascertain environments for frequency ranges of interest across the entire vehicle. Additionally, the aerosciences community lacks the experimental data for further development and validation of CFD models. Therefore, wind tunnel testing is the preferred and best method for characterizing the unsteady, transonic flow environments and area-integrated loads (Steva, Pollard, Herron, & Crosby, 2019; Piatak, Sekula, & Rausch, 2016). Experimentally, the current state-of-the-art is using sophisticated, miniature unsteady pressure transducers. These sensors have been used for many decades given their accuracy and simplicity to operate. However, since they are discrete sensors but are employed to measure a complex spatial phenomenon, a large sensor count is required to understand the dynamic environment over an entire vehicle. Furthermore, the sensors traditionally have long lead times, are relatively expensive given the required high sensor count, and are usually too delicate to be repurposed beyond their initial installation.

The continued absence of accurate and efficient physics-based CFD modeling tools to predict launch vehicle fluctuating pressure environments hinders innovation due to the cost and schedule impacts required to qualify new configuration concepts using only ground and flight tests.

However, to mature accurate and efficient computational modeling tool, high spatiotemporal surface pressure measurements from ground test experiments are required.

Because of these challenges, the NESC invested in the research and development of innovative technologies (Winski, et al., 2019; Schuster, et al., 2016) that would further develop the aerosciences community's understanding to predict and manage unsteady, separated flow. Several of these new technologies are optical and offer increased spatial resolution for measuring the complexity of unsteady flow (Salim, Uram, Childers, Vishwanath, & Papka, 2019).

In all fields of science, the scientific frontier is studying the evolution of complex phenomena in space and time. Given the cost continues to decrease for sophisticated, high-throughput technologies, like detectors, imagers, and sensors, the scientific research communities are using powerful optical methods to enable scientific discovery at higher resolutions than previously available.

In medicine, using X-ray technology, researchers at Argonne National Lab and University of Chicago successfully imaged an entire mouse brain, approximately one cubic centimeter in volume, at one micron resolution, producing two TB of data in six hours, producing one of the fastest approaches for whole brain imaging and revealing new details about the brain's structure (Foxley, et al., 2021)

In astronomy and astrophysics, NASA's James Webb Space Telescope uses three scientific camera and a spectrograph to collect sharper and higher resolution images of deep space than its predecessor, Hubble Space Telescope, generating approximately 57 GB of data each day, approximately 20 TB each year and almost one million miles away from Earth (Gray, 2022). The Vera Rubin Observatory, a joint venture between the National Science Foundation and Department of Energy, is currently being built in northern Chile. Its 3.2 Gigapixel camera, the largest digital camera ever constructed for ground based optical astronomy, will collect 20 TB of data each night for 10 years, approximately 500 PB total (Vera C. Rubin Observatory, 2023).

In technology, Waymo's Chrysler Pacifica driverless vehicles have 19 cameras, millimeter wave radar, and LiDAR sensors coupled with machine learning algorithms working together to determine its surroundings and detect the speed and distance of nearby objects, generating approximate 40 TB of data collected during an eight-hour operation (Waymo, 2023).

In aerospace, Korthäuer et al. (2023) combined two optical technologies, a high-speed background-oriented Schlieren measurement with stereo high-speed model deformation cameras, to investigate the complex phenomena of self-excited fluid-structure interactions on an elastically suspended supercritical airfoil. Garbeff et al. (2019) deployed dual high-speed and low-speed

cameras for acquiring time-resolved shadowgraph images and a mid-wave infrared (IR) camera with 6 millisecond exposure times to provide qualitative measurements of the near field turbulent fluctuations around a scaled launch vehicle wind tunnel model.

Data is essential for scientific discovery. Science advances by collecting data to replace previous models with new models that better explain the observation (Fortunato, et al., 2018). The data volumes and velocities from high-throughput technologies are increasing, pushing past the current computational resources within the ground testing facility to the point where additional computational resources are required. But big data challenge has left us “data rich, and analysis poor” (Clery & Voss, 2005) due to lack of computational power and storage within the ground testing facilities’ embedded system environment.

Advanced infrastructure is required to increase knowledge creation and harness the growing data volumes and velocities. The technology used in these facilities is changing; therefore, the processes, algorithms, and hardware used to move and analyze data must also change. The local computation and storage needs of the experimental communities have grown beyond what is available (Bear, Parkinson, Gerhardt, Sibony, & Enders, 2021) to the point that high-performance computing facilities are not just desired, but by many, required. To fulfill this computational need, ground testing facilities could advocate for increased funding to purchase additional computational resources and personnel to manage and maintain this consumable equipment. However, even if resources were available, many facilities lack the floor space and utilities (e.g., power, cooling, network infrastructure) to host these machines (Bear, Parkinson, Gerhardt, Sibony, & Enders, 2021).

It is not by chance that the Department of Energy, Department of Defense, and NASA own and maintain clusters of computers to rapidly perform complex calculations, which for this article will be referred to as high-performance computing (HPC) and includes cloud and edge computing resources. Each of these federal department operates on the frontier of science in fluid dynamics, material science, astronomy, and other sub-disciplines to understand the complex, chaotic evolution concerning the flow and interaction of particles, which usually produce large data sets in space and time. Oftentimes, flagship HPC resources reside within a mile of the experimental laboratory and should be leveraged to maximize the computational resources available. In June 2022, the Department of Energy's Oak Ridge National Lab brought online Frontier (Oak Ridge National Lab, 2022), the first exascale (10^{18}) and top-ranked supercomputer in the world, with 1,102 petaflops/sec, or 1.102 exaflops/sec, almost three times more than the second-ranked supercomputer in the world. The Department of Energy and NASA, respectively, have 10 and three

independent supercomputing facilities in the top 100 facilities in the world (TOP500, 2022). This confirms that the complex challenges being surmounted on the frontiers of science, either experimentally or through simulations, will be met with the machinery and computational environment required.

Not every ground testing facility will need the sophistication of exascale HPC, but Abeykoon et al. (2019) highlight that a facility should be paired with the appropriate computation resources. There is not a one-size-fits-all solution and depends on what computational resources may be locally available, the computational demands, and complexity of the problem (Elias, Chard, Libera, Foster, & Chaudhuri, 2020). Working in a HPC environment is foreign to many experimental scientists and researchers and will require further development by the HPC community to create an environment where researchers are not met with obstacles. Due to a disruption in the current experimental and computational facilities workflows, Salim et al. (2019) stated HPC facilities are generally underused by the experimental community.

Nevertheless, growing data volumes and velocities and a corresponding need for more computational power and storage is a common theme among experimental facilities in all scientific fields. Table 1 lists several premier scientific facilities from across a broad range of scientific domains and their respective data rate per experiment or per year. The large data rates listed confirm the trend that data volumes and velocities will only continue to grow when studying the frontiers of science. Increased computational power is necessary for scientific discovery. High-performance computing began after World War I, with the need for more computational resources required for the design of weapons and aircraft (Senning, 2020).

With the increase in computational power in the 1980's and 1990's, the aerospace industry transformed how advanced aerospace vehicles were designed. The Boeing 777 was the first made from simulation software (Norris & Wagner, 1999). Now, there is a stronger emphasis on computer simulation and less time and money on flight testing.

In medicine, Gorzynski et al. (2022) were able to decrease the time to sequence human DNA within hours, rather than weeks by using distributed cloud-based computing and set the first Guinness World Record for the fastest DNA sequencing technique by employing more powerful computing resources. The additional of increased computational resources, enables physicians to rapidly diagnose patients knowing if their illness is viral or rooted in their DNA.

Discovery of the Higgs boson particle was only made possible by using European Council for Nuclear Research's (CERN) Worldwide LHC Computing Grid and U.S. Department of Energy's Oak Ridge National Lab's Titan supercomputer given the 150 million sensors within in the ATLAS

Table 3: Data Rates Within Large, Government-Owned Scientific User Facilities

Instrument	Field of Science	Experimental Facility	Computational Facility	Data Rate	Reference
Particle Accelerator	Particle Physics	CERN Large Hydron Collider (LHC)	CERN Data Centre Worldwide LHC Computing Grid	25 GB /s	(Rao, 2020)
Particle Accelerator	Particle Physics	Linac Coherent Light Source-II SLAC Linear Accelerator Lab	National Energy Research Scientific Computing Center (NERSC) Lawrence Berkeley National Lab (LBNL)	1-5 GB / s	(Thayer et al., 2019)
Flame Spray Pyrolysis	Nanomaterial manufacturing	Combustion Synthesis Research Facility Argonne National Lab (ANL)	HPC-GPU Enhanced Cluster at ANL	1 TB / experiment	(Levental, et al., 2020)
X-ray Beamline	Material and Biological Science; Physics; Chemistry	Advanced Photo Source (APS) ANL	Argonne Leadership Computing Facility (ALCF)	~2-12 GB / s	(Salim, Uram, Childers, Vishwanath, & Papka, 2019) (Rao, 2020)
UV / X-ray Beamline	Material and Biological Science; Physics; Chemistry	Advanced Light Source (ALS) LBNL	NERSC LBNL	200 PB / year	(Department of Energy, 2022)
Wind Tunnel	Aerosciences	NASA Ames Unitary Plan Wind Tunnel	NASA Advanced Supercomputer (NAS)	2 – 5 GB / s	(Roozeboom N., Baerny, Murakami, Ngo, & Powell, 2020)

detector (Belyaev, et al., 2017). As highlighted with these examples, once the data is in suitable computational environment, much more innovation can be applied, and new tools can be developed and leveraged to allow for broader, more accurate calculations (Brunton, et al., 2021).

For Ground Testing 3.0 to be realized, automation is essential. Ground testing 2.0 was powered by automating data acquisition and data recording tasks that were highly manual. This automation led us to acquiring more data and realizing the advancements we have seen in the design and development of advanced aerospace vehicles since. The art of data collection has been honed at these facilities over the decades since. Data acquisition automation has been well established and enabled by custom and off-the-shelf software in these facilities. However, the automated process of extracting knowledge from the automation of scientific data collection has not kept pace (Schmidt & Lipson, 2009). Now with the power of artificial intelligence, more of the scientific workflow can be automated (Wozniak & Schwarz, 2019). With the deluge of data, humans cannot possibly look at or derive conclusions from the growing data sets produced by the automated data acquisition at ground testing facilities. The growing fields of artificial intelligence (AI) and data-driven optimization can be used in making informed, traceable, automated decisions to accomplish data analysis tasks.

Automation enables more ambitions and intricate scientific challenges to be explored since a greater number of parameters can be examined. The experimental design and decision making process of the measurements mostly remains determined by individual researchers (Noack & Zwart, 2019). Today, researchers combine existing knowledge from previous experiments or experience with the most recently collected measurements to adapt their test matrix to acquire the best data most efficiently with respect to time and money. The manual process is time consuming and becoming increasingly unmanageable. There is substantial drain on the human capital and when the parameter space is larger than a human can consume, biases are introduced.

To avoid these biases, it is critical that more of the scientific workflow be automated. This includes using data-driven optimization and machine learning tools. Coutant et al. (2019) demonstrated that hypothesis-led experiments improved their systems biology model to a greater extent than achieve using high-throughput experiments. With this finding, the authors concluded more ideas from AI will be required to improve performance and giving the AI tools the ability to communicate goals and intentions to human scientist fundamental enhancements would be achieved.

New developments in machine learning make it possible to extract more from this data.

With moving the experimental, scientific data to the right computational environment, new

innovations will accelerate scientific discovery. Taira et al. (2017), Brunton et al. (2020), and Brunton et al. (2021) give thorough reviews of modal, machine learning, data-driven analysis used in fluid mechanics and the aerospace industry and are highly influential resources for learning more in-depth about the work being conducted in the aerospace industry with respect to these analysis techniques. The uses of machine learning and data-driven optimization in the aerospace industry are broad. In the spirit of ‘necessity breeds innovation’, this paper is not intended to outline the perfect analysis tools or a single solution, but rather give need and motivation to accelerate the development and deployment of such tools. The general goal is to determine the relationship between response variable and input variables using a wide-range of solutions. Sparse and symbolic regression, Bayesian optimization, reduced order modeling, response surface methodology, regularization and genetic algorithms are common algorithms used in the aerospace industry.

A few recent examples of data-driven or machine learning algorithms developed and used in the aerospace industry are given here. Using data from a numerical simulation of flow past a flexible wing, Hickner et al. (2023) developed a low-order linear model to trace flight maneuvers with model predictive control using a system identification approach. Sotoudeh et al. (2023) implement the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm (Brunton, Proctor, & Kutz, 2016) to analyze gust response in low Reynolds number conditions. Loiseau and Brunton (2018) combined the SINDy with proper orthogonal decomposition (POD) to identify sparse reduced-order models of several flow systems. Li et al. (2022) used Dynamic Mode Decomposition (DMD), a dimensionality reduction algorithm, to process high spatiotemporal surface pressure measurements on a wind tunnel model using unsteady Pressure-Sensitive Paint for effective diagnosis of unsteady, aerodynamic phenomena. Choppala et al. (2023) developed a supervised machine learning algorithm using random forest regression to determine the optimal sensor location for measuring stress on a plate. Sotoudeh and Eger (2023) used machine learning to refine mesh resolution when needed. Mesh resolution is an important cornerstone of CFD, finite element analysis (FEA), and optical methods. Kirenga, Lee, and Zolotas (2023) discussed their development work to develop a low-cost solution for monitoring rail yards using drones and imagers by implementing an object detection algorithm to increase efficiency and decrease cost. Sun and Ma (2023) also used drones and imagers combined with reinforcement learning for automated aircraft inspection to increase efficiency and safety over current human-based inspections. Lastly, Kefalas et al. (2022) employed symbolic regression on high frequency, exhaust gas temperature data from an actual flight to reveal mathematical relationships between the exhaust gas temperatures and

engine parameters. Knowledge about what does and does not matter can improve resource allocation (Roth & Singhal, 2022).

Machine learning is not a blackbox, rather, it is a growing set of optimization and regression techniques to build models from data. Human intelligence is essential in each step of the development. For the aerospace industry, these models will be physics-based and must be generalizable, explainable, and certifiable for safety-critical applications (Brunton, et al., 2021). Moreover, machine learning algorithms and optical methods in aerosciences are complementary since much of machine learning is focused on imaging science. There is not one data-driven optimization or ML algorithm or category of algorithms, but rather this highlights a rich area of development for hybrid methods combining first principles and data-driven optimization algorithms. Data-driven optimization and machine learning tools have become another tool in the aerospace design toolbox and will continue to improve as we develop and deploy more methods collectively (Jordan & Mitchell, 2015)

Several ground testing facilities within the U.S. Department of Energy and NASA are actively pursuing near-real-time and autonomous steering of experiments (Noack & Zwart, 2019; Bicer, et al., 2020; Levental, et al., 2020; Elias, Chard, Libera, Foster, & Chaudhuri, 2020). The numerous sensors installed in the ground testing facility and test article give new opportunities for monitoring and guiding scientific experiments.

At the DoE's Argonne National Laboratory's Materials Engineering Research Facility using Flame Spray Pyrolysis (FSP), a process that produces a wide range of nanomaterials and capable of producing one TB per experiment, Levental et al. (2020) outline a hybrid CPU-GPU Difference of Gaussian method to extract the volume of unburnt solution and realize near-real-time steering of FSP experiments. Elias et al. (2020) designed a machine learning platform to optimize the ground test facility through an AI-automated process for real-time steering of experiments by using local computational resources for speedy response for control and HPC resources for large-scale analysis.

Researchers at DoE's Lawrence Berkeley Laboratory, use AI and ML tools to automate experiments at several X-ray scattering and synchrotron infra-red microscopy laboratories to maximize instrument usage, reduce bias, and decrease risk. Noack and Zwart (2019) construct a surrogate model, using Gaussian Process Regression (Rasmussen & Williams, 2006), to autonomously direct an experiment.

Bicer et al. (2020) define a high-performance system allows for testing parameters to be adjusted dynamically during an experiment by using phantom and real-world optical tomography dataset for X-ray experiments at the DoE's Argonne National Lab.

There are more than 50 premier ground test and scientific user facilities across the federal government, and many of these facilities have limited availability requiring customers to book test one to two years in advance. To improve how these facilities serve the scientific community, advanced experimental techniques like autonomous and near-real-time steering of experiments are required.

To solve this sort of niche challenge on increasing knowledge creation, analogical reasoning—taking something familiar and making it new or vice versa—is a powerful tool (Gentner & Smith, 2012). Before the concept of gravity and physical forces existed, German astronomer Johannes Kepler, credited as the inventor of astrophysics, used a series of analogies from heat, odor, light, magnetism, fluids, and optics to propose that planets travel elliptical paths around the sun based on their distance (Gentner, et al., 1997). Epstein (2019) highlights that the genetic algorithms used in the field of optimization are inspired by Charles Darwin's theory of natural selection. Romano (2009) used fluid dynamics, specifically the Reynolds number equation, as an analogy to examine the relationship between a supply chains structure and performance to demonstrate the properties of a fluid flowing in a pipe can be associated to the characteristics of products flowing in a supply chain network.

There have been few instances where supply chain models have been used in problem solving in the aerospace industry. Bird and Langelaan (2020) used a multi-armed bandit model to determine the optimal altitude for small, unmanned aircraft systems (UAS). Roy and Tomlin (2008) used a supply chain model as a computationally tractable estimation for the more computationally intensive nonlinear partial differential equation model for traffic flow management.

Aryal et al. (2018) highlighted innovative technologies that would force a change in supply chain management. The inverse is also true. By mapping the new challenges limiting the pace of scientific discovery in ground testing facilities to well-studied supply chain management models developed over the last 100 years (Luthans, Luthans, & Luthans, 2015), we have the wide variety of tools required to solve these challenges and achieve the goal of increasing knowledge creation.

The movement of data from the manufacture (i.e., wind tunnel) to the customer is analogous to the movement of a material good in a supply chain. Within supply chain management research, one area that has formed in recent years is the data supply chain. Spanaki et al. (2018) propose the data supply chain and build on theoretical and empirical foundations of the traditional manufacturing

supply chain where data is the primary artifact in the supply chain. Scientific data is a definitive subcategory of data to be considered and studied. The term "scientific data supply chain" should be construed as a paradigm for the management, distribution, and procurement of processed scientific data commencing with the acquisition of raw scientific data.

In this context, data is the material moving through the supply chain. A unit of data, in its smallest form, is a bit. Bits require infrastructure to collect, store, and transfer. Like the material supply chain, the scientific data supply chain will realize efficiencies in costs when the scientific data generated during an experimental test campaign is viewed as a product to be managed. Connecting all parts of the scientific data supply chain will influence the whole system (Parkinson, Krishnan, Ushizima, Henderson, & Cholia, 2020).

But still it remains, an abstract model for knowledge creation that highlights why to move from current operations to next-generation, data-driven operations would be advantageous, especially in the infusion of digital transformation into the standard operations of these sophisticated experimental facilities. One of the strengths of the supply chain management field is the broad application of supply chain management methods is one of the strengths of the field (Jónasson, Ramdas, & Sungu, 2022). For increasing knowledge creation within scientific research, the multi-armed bandit and the multi-period newsvendor models are excellent models to apply. The multi-armed bandit framework, formalized by Robbins (1952), is useful for sequential decision-making problems between competing choices in which a limited number of resources must be allocated. Each choice has an unknown reward distribution, and the user desires the highest possible reward. Time and budget limit the number of chances to "pull the arm". Therefore, the multi-armed bandit's intuitive framework, robust performance, and ability to learn from less feedback (Bouneffouf, Rish, & Aggarwal, 2020) produces a powerful framework for business operations (Schwartz, Bradlow, & Fader, 2017). The user must explore the choices to realize the reward distribution. Once a desirable solution is identified, the user can exploit this option. The multi-armed bandit is often referred to as the "exploration versus exploitation" dilemma.

Another factor that impedes knowledge creation is the inability to quickly share data. Spanaki et al. (2018) designed the data supply chain to guarantee timely data sharing among its users. The optimal planning in supply chains (SC), regardless of what is going to be managed, could be determined when all information about customers' demand is known. However, when it comes to stochastic environments (Porteus, 1990), planning confronts more difficulties. The customers of scientific user facilities come from the public and private sectors and have stochastic demand. In a stochastic demand environment, the demand for a product or item is not known with certainty, and

it may vary from one period to another. This uncertainty in demand can lead to overstocking or stockouts if inventory levels are not properly managed, resulting in some undesirable outcomes (Chiu & Choi, 2016). Overstocking and understocking in the setting of scientific user facilities are translated into how many more or fewer experiments have been done, respectively. The former imposes a remarkable financial burden while the latter limits the pace of scientific discovery. In the literature on supply chains, one promising strategy for confronting the complexity of the stochastic nature of customer demand is formulating the problem through the idea of the Multi-Period Newsvendor problem. The multi-period newsvendor problem is a classical stochastic inventory control problem with a long history in literature. First introduced by (Arrow, Harris, & Marschak, 1951), the multi-period newsvendor works by considering historical demand data to forecast future demand. In this direction, the retailer can make informed decisions on how much inventory to order in each period to meet customer demand while minimizing costs (Matsuyama, 2006). observing and tuning inventory levels based on demand patterns, The retailer can circumvent overstocking and understocking by continuously observing and tuning inventory levels based on demand data, allowing the retailer to better serve its customers (Qin, Wang, Vakharia, Chen, & Seref, 2011).

Considering the complexity and multi-disciplined nature of the scientific data supply chain problem, the multi-period newsvendor (Matsuyama, 2006) model is valuable because it emphasizes the ability to share information and adjust demand more frequently, and therefore effectively increasing the number of ordering periods during a single selling season, producing higher rewards for the retailer and overall supply (Wei & Chen, 2021).

To this end, through hybridizing the multi-period newsvendor model and multi-armed bandit framework, we proposed a hybrid approach, the Multi-Armed Newsvendor Model to take the benefit of both in the regard of accelerating knowledge creation. The multi-period newsvendor model emphasizes the importance of information sharing, while the multi-armed bandit framework instructs retailers on how to best deal with the exploration vs exploitation dilemma. Exploration is expensive while testing at a scientific user facility. The customer has a finite set of budget and time allotted to test their hypotheses. The multi-armed newsvendor model guides the customer in exploring in a cost-effective measure while gathering enough information to prove or disprove their hypotheses. It is imperative experimental data is shared among the cross-discipline team while actively testing to heuristically make informed decisions.

Ground test facility customers are attempting to solve their exploration versus exploitation problem and desire the maximum amount of knowledge at minimum cost. Fortunately, the fields of operations and supply chain management know what good processes look like and how to

efficiently design a system to solve the exploration versus exploitation dilemma. As a not-for-profit operation, the ground test facility is not responsible for each customer's ability to maximize science at the lowest cost, however, the facility is responsible for providing a system that allows their customer to reap the maximum reward by changing the rules of the games by shortening the time windows between decision making improves the process and by leveraging available resources and employing HPC, which allows the customer to reoptimize faster and, consequently, obtain better solutions. Depending on the test campaign objectives, data volume and velocity, and availability of computational resources, the exact solution will vary.

Methodology and Solution Approach to Transform the Scientific Data Supply Chain

It would be advantageous to examine how the data delivery at ground testing facilities is organized. To aid our discussion and confirm a customer has advanced our understanding while testing within a ground testing facility, we will use the Data-Information-Knowledge-Wisdom (DIKW) pyramid framework from the information science field to explain the development of human understanding. Consulting the DIKW pyramid in Figure 9, raw data plus context equals information, both are objective in nature. Adding subjective expertise, academic, professional, or personal, the knowledge echelon is achieved. When a team or individual is able to take an action based on that knowledge, we have progressed to the state of wisdom. Each step adds value and meaning to the process.

Figure 10 gives a generalize depiction of the scientific data supply chain and timeline for how it is organized today at many ground testing facilities. Test planning is generally accomplished over a year. Requirements are gathered from the customer, and a test campaign is planned. The model hardware and instrumentation are designed, sourced, and assembled. Data-processing models and a test matrix are generated. Finally, the test article is installed in a scientific user facility, and the test campaign begins. As highlighted by Bear et al. (2021), the ground testing facility is an embedded system, shown by the bold line Figure 10, and is severely spatially limited since it hosts many functions. Within the scientific user facility, the facility control system, facility crew (approximately 10 people), instrumentation suites (Instr. Suite), local computational resources, and a limited number of customers are all hosted inside the physical walls of the experimental facility. The facility control system is the main system that automates and controls the operation of the facility. The facility crew will have a broad range of the required experts: test engineers, electricians, mechanics, instrumentation technicians, and owners of various instrumentation suites. After the test campaign ends, processed results are mailed to the customer on an encrypted external hard drive, and detailed analysis can begin two to four weeks post-test. If a customer is unable to

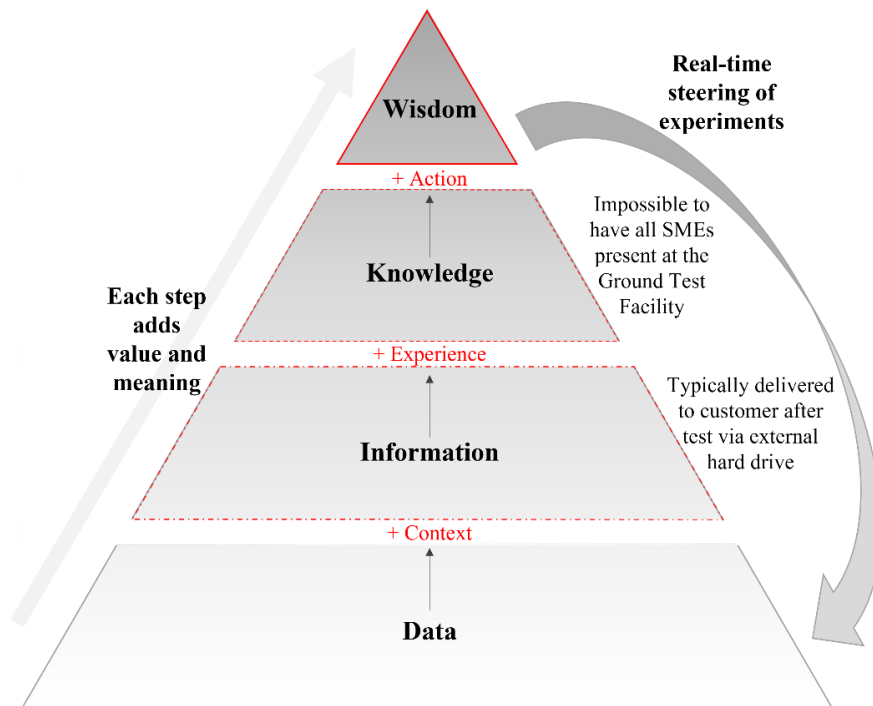


Figure 9: Data Information Knowledge Wisdom (DIKW) Framework

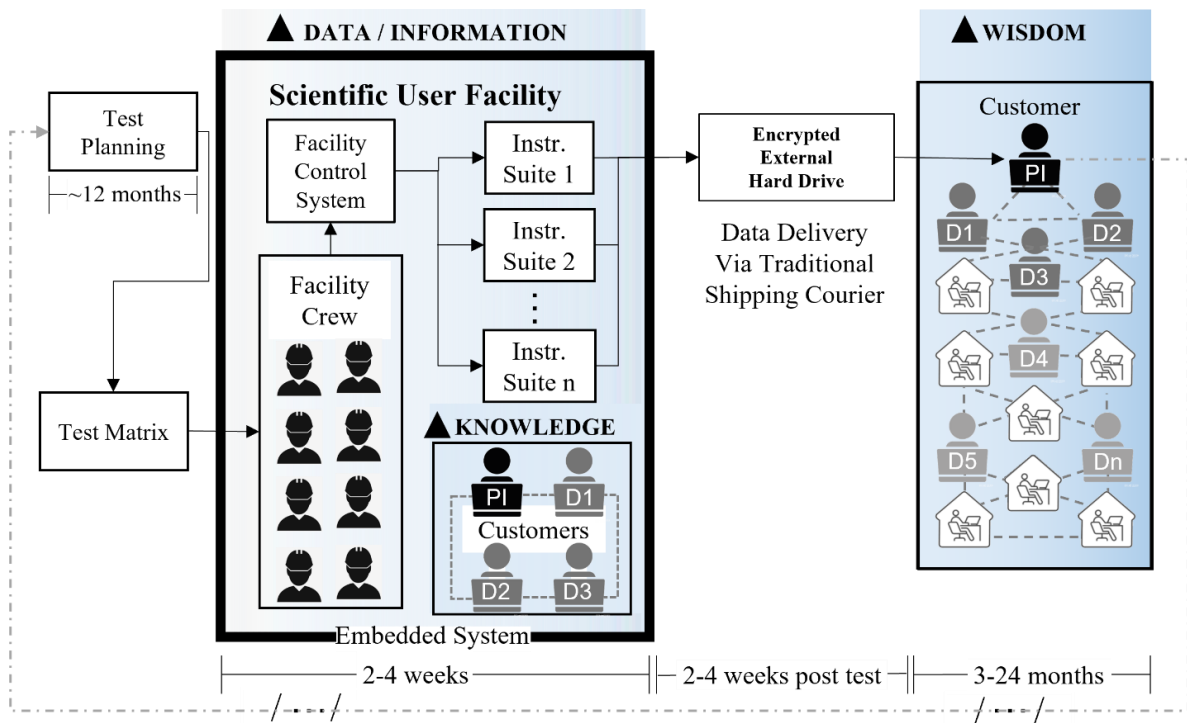


Figure 10: Traditional Scientific Data Supply Chain

prove or disprove all hypotheses, the test campaign must be re-ran, and with the limited availability at large-scale ground testing facilities, delaying additional testing by a year or more. As shown in Figure 10, data and information are acquired and available within the embedded system. For the limited number of customers physically present at the ground testing facility, some level of knowledge is created, but their ability is limited to the individuals and their subjective expertise. Wisdom (i.e., actionable knowledge) is not truly realized until several weeks after the test when a larger group of individuals has access to the experimental scientific data acquired during the test campaign.

The challenges faced by ground test facilities in terms of their ineffectiveness in exploiting the full power of data for knowledge creation, give reason to find a solution. By examining how the scientific data supply chain is structured, we have revealed areas where ground testing facility can most effectively and efficiently innovate (Carter, Rogers, & Choi, 2015). The length of the data supply chain in Figure 10, is an indicator of the vulnerability posed to the customer and the ground testing facility. Longer supply chains are more vulnerable to disruption and longer recovery response times (Graves et al., 2022). Since data is the artifact moving through the supply chain, more attention should be dedicated towards decreasing the lead time of data being delivered.

The data must be moved faster from the facility to a larger number of customers. In supply chain management, one of the most significant factors in determining the transportation method of a product is the value-to-weight ratio. Products with a high value-to-weight ratio should be moved as quickly as possible. Given the cost (\$1-2 Million per test campaign) and the time to plan (approximately one year), data should be moved as quickly as possible. Electronic delivery through a data pipeline stands alone as the only method worthy of the data's value-to-weight ratio. Processing and sharing data using HPC resources (including edge and cloud computing), will directly increase the number of customers able to access the data in near-real-time. By building the data pipeline, we are converting a high volume of data at a fast rate from data to knowledge, and ideally data to wisdom.

However, simply moving the data faster from one location to another only solves the transportation challenges. Moving the data to a location with accessibility to more individuals in near-real-time with enterprise-level tools for sharing and data analysis exposes the value of information sharing in a supply chain. It is the benefits of information sharing that give the motivation to improve the transportation method. Information sharing is essential for the ground test facility to achieve its goal in creating new knowledge and accelerating scientific discovery. In a knowledge-service setting, "productivity is measured as the amount of information exchanged

per unit time” (Jónasson, Ramdas, & Sungu, 2022). Research by Li et al. (2016) and Islam et al. (2022) concludes providing each stage with timely information, variability, uncertainty, and lead time lessens across the supply chain producing higher profits (knowledge creation). The key finding from the Department of Energy Accelerating Scientific Knowledge Discovery Working Group report (Department of Energy, 2022) was to accelerate scientific discovery, data must be process and shared among customers as quickly as collected.

The U.S. government, as the owner of the ground testing facilities, highly desires customers to maximize their reward while testing by proving or disproving their hypothesis and ultimately use the knowledge extracted to stimulate economic growth. This gives the motivation to address the transportation method and digitally transform the scientific data supply chain to reap the rewards by sharing the information while a customer is testing at a ground testing facilities, shown in Figure 11.

Figure 11 shows a digitally transformed ground test facility where the constraints of the embedded system have been removed and a data pipeline moves the experimental data / information to a HPC environment giving access to a larger number of customers regardless of physical location allowing for knowledge to be created. The HPC environment also includes computational hardware capable of harnessing the power of experimental data through data-driven analytics and access to existing dataset, like CFD-simulated data. Combining resources available within an HPC environment and the expertise of a larger group of individuals and teams permits autonomous data-driven decisions to be made during the test campaign.

The NASA Ames Research Center’s (ARC) Unitary Plan Wind Tunnel’s (UPWT) 11-by 11-foot transonic test section (NASA, 2022) is well-known for its optical accessibility and optical measurements in the aerospace ground testing community. With the increase in optical technologies deployed at the facility, the UPWT has become a big data generator with insufficient computational resources within its current capabilities. During previous demonstrations where data volumes greater than 1 TB were generated, data was loaded onto encrypted hard drives and moved to an HPC environment weeks or months later. These demonstration test concluded an increase in computational power, tools, and storage was needed (Roozeboom, et al., 2018).

In 2019, NASA conducted a first-of-its-kind demonstration called Project: Red Rover in ground testing by using an HPC facility to process and share experimental results in near-real time (NASA, 2019). Through this effort, the scientific data stream from an experimental facility, ARC UPWT, was transferred to a computational facility, NASA Advanced Supercomputer (NASA, 2022), both

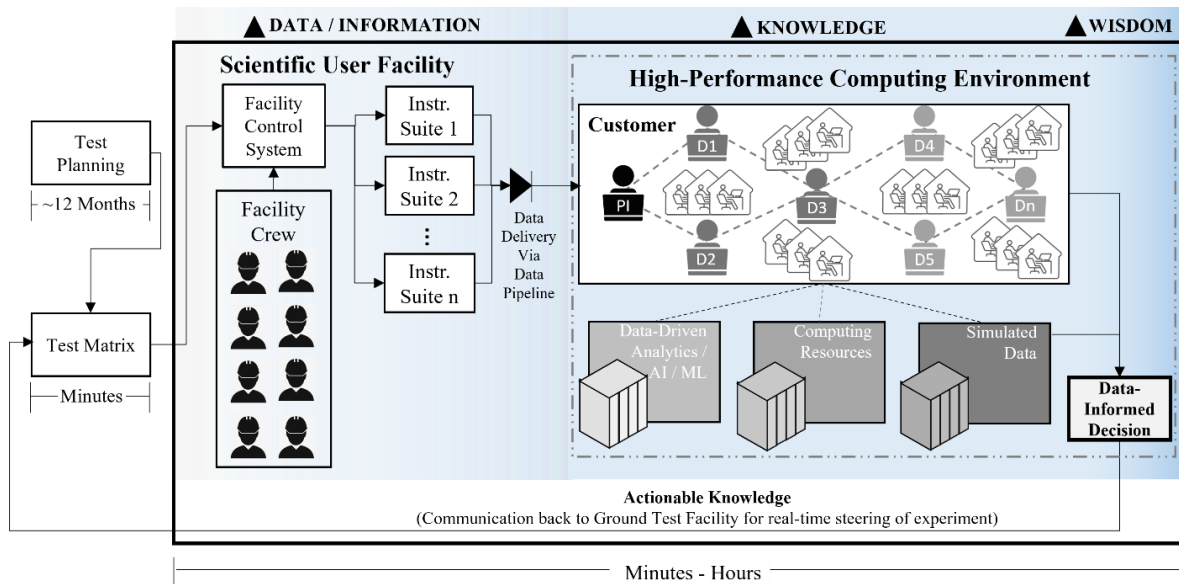


Figure 11: Transformational Scientific Data Supply Chain

located at NASA Ames Research Center (ARC), Moffett Field, CA. During the Project: Red Rover demonstration, 150 TB of data, collected over four shifts of operations were securely transferred between the experimental and computational facilities, processed on the supercomputer, and distributed to local and remote customers in near-real-time during an active wind tunnel test campaign. This demonstration would mark the beginning of a new era of ground testing where hypotheses are driven by data (Brunton, et al., 2021), experimental facilities will begin to leverage HPC resources for near-real time analysis and customers and facilities are empowered to make decisions.

The U.S. government, as the owner of the ground testing facilities, highly desires customers to maximize their reward while testing by proving or disproving their hypotheses and ultimately use the knowledge extracted to stimulate economic growth. This gives the motivation to address the transportation method and digitally transform the scientific data supply chain at ground testing facilities.

Ultimately, it is not the basic technology of connecting experimental and computational facilities, it's the major breakthroughs in data-transfer, processing, sharing, and harnessing that will change how the aerospace industry designed advanced aerospace vehicles.

Use Case

The prediction, measurement, and management of unsteady flow is the most pressing challenge facing the field of aerosciences and virtually affects the entire discipline. Fluctuating pressures on and near the vehicle's surface can induce extreme dynamic loads and vibrations and are severe enough to cause the loss of the mission. These fluctuating pressure levels are caused by high-speed flows originating from flow separation, flow reattachment, vortex shedding, transonic shock oscillations, and compression/expansion corner geometry (Piatak, Sekula, & Rausch, 2016).

Launch vehicles encounter unsteady pressure loads in the form of 1) buffet and 2) aero-acoustic noise. Buffet is defined as a repeat loading of low-frequency random pressure fluctuations during launch vehicle (Cole, Erickson, & Rainey, 1970) and is applied as an area-integrated time-varying force at specific longitudinal stations. Aero-acoustic noise are at higher frequencies and more localized.

Typically, the buffet environment is most extreme at conditions in the transonic regime and capable of exciting the vehicle dynamic vibration modes. This vibroacoustic loading can cause mechanical and electronic malfunction, fatigue failure, payload vibration, and unfavorable occupant environments. Therefore, rigorous vibration qualification testing is imposed on parts and components of the vehicle to avoid failure during launch. In addition, loads from buffet are one of

many environments contributing to the total loads of a launch vehicle during launch and applied as part of a coupled loads analysis to determine the vehicle's structural response and safety margins (Piatak, Sekula, & Rausch, 2016) .

Performing a large-scale wind tunnel test has produced a sufficient method for characterizing both the vehicle buffet and aero-acoustic environments. Empirical methods are available for predicting external fluctuating pressure levels produced while the launch vehicle ascent, however the uncertainty of empirical methods produces undesirable under-conservative or over-conservative engineered designs (Steva, Pollard, Herron, & Crosby, 2019)

For NASA's Space Launch System, the vehicle continued to evolve over the years and several changes were made between Block 1, Block 1B, and Block 2. Given that buffet and aero-acoustics environments are closely linked to details of the model's geometry and specific to a given, this presented yet another challenge to the SLS program to define unsteady environments for a vehicle that continued to evolve. It is imperative that predictions for fluctuation loads be made available to the program early in the design. Early testing dictates design. Analysis needs to have the appropriate amount of conservatism, too much or too little will greatly affect. Vibration level to be imposed during such testing usually must be specified early in the design program. If the vibration level is too high, then great expense and efforts are exhausted to meet overly conservative requirements. On the other hand, if vibration levels are too low, then redesign efforts are required downstream, disrupting cost and schedule.

NASA began developing the Space Launch System (SLS) launch vehicle in 2011. In 2012, NASA began testing the Space Launch System's (SLS) performance against unsteady, chaotic flow. After the initial wind tunnel tests in 2012, engineers produced results estimating the aerodynamic load, introduced by the chaotic, unsteady environment the SLS rocket would fly through, would not meet program requirements. Understandably, this caused a significant concern that an extensive redesign would be required. A special task team was assembled to address why these results were found and identify options to mitigate high aerodynamics load estimates. This included examining the models used to estimate the aerodynamic loads from the original test and examine new hardware to mitigate the aerodynamic loads (Sekula, Piatak, Rausch, Florance, & Ramey, 2015). Fortunately, these mitigation options would be assessed eight months later, in 2013 during a previously scheduled wind tunnel test, and then again six months later, in 2014, when the SLS Program requested high priority access to the scientific user facility. During these tests, additional instrumentation was added to increase the sensor's spatial resolution measuring the unsteady surface pressure fluctuations, and new numerical models would be produced based on the

higher resolution data. Both solutions were expensive options to NASA. For this effort alone, the approximate cost was \$2 Million for additional testing and \$1.5 Million for the special task team. Neither of these includes any cost associated with schedule delays, additional workload imposed to NASA and its commercial partners, or increase travel requirements.

Solving challenges associated with unsteady flow requires a multi-fidelity and multi-disciplined approach. There are many tools deployed in the aerospace design community. While the MANV framework assumes the reward distribution is unknown, once you have reached the readiness level to test at a large ground test facility, a customer will have several computational and small-scale datasets. This information will aid the customer in their exploration versus exploitation to prepare the customer to use time and money most effectively by rapidly converting information to knowledge. Quantity and quality of data is growing, and we should be utilizing all the tools. CFD enables us to run and gather experiments more intelligently. We must continue to develop algorithms that convert data into actionable knowledge. Here are several examples how to effectively address the exploration vs exploitation dilemma when assessing unsteady flows.

Example 1:

Explore: Given that the state of the art for CFD is prediction of steady, attached flow using Reynolds-Averaged Navier-Stokes (RANS), low-fidelity, steady-state CFD accurately predicts where large fluctuations are expected and will prescribe placement of the limited discrete sensors before the ground tests planning begins.

Exploit: Then execute high-fidelity experiments concentrate resources on areas of large fluctuations and conditions of interest

As the program matures, vehicle configurations will evolve and change. It is impossible to retest in a wind tunnel or run time-accurate CFD cases for every configuration and condition immediately. Example 1 is an effected solution to continue moving forward with design until a testing facility is available.

Example 2:

Explore: Before a ground test, run time-resolved CFD cases at conditional critical for vehicle design based on historical data from similar geometries.

Exploit: At ground test, collect time-resolved, high spatiotemporal data with optical technologies at same time-resolved CFD conditions. At this point, exploit the fact that it is more effective to generate data at ground test facility. Depending on the sensitivity of a computational model, vary those parameters within the ground test. Use data-model fusion to drive what testing conditions are needed for comparison to the CFD data. Using optical

technologies early in the design process will generate valuable datasets with desirable high spatiotemporal resolution and will limit or avoid over-conservative or under-conservative estimates due to limited resolution of data generated.

- Doing this early in the test campaign allows for in time-consuming comparisons to be made between experimental and computational data sets. Differences between the two datasets are expected.

Example 3

Explore: Given the MANV solution strategy, next explore broadly among the design space. Use data-driven, statistical tools like response surfaces, sparse regression, dynamic model decomposition to understand relationship between dependent and independent parameters.

- **Exploit:** In areas where statistical models exhibit sensitivities, exploit by collecting data at conditions of interest.

As discussed with the SLS use case, one portion of a ground test was to explore how the addition of fences mounted on the solid rocket boosters or the launch vehicles core around the attachment hardware would influence the dynamic load on the vehicle. During this test, engineers used a handful of sensors located near the region of interest to assess how the fence geometry affected the dynamic loads. Given their relatively small size, engineers would uninstall the fences and by using hand tools slightly change the fence geometry. The model pieces were reinstalled and the test resumed. As highlighted previously, protuberances and geometric changes have a large affect on the localized dynamic loads. By digitally transforming the scientific data supply chain and bringing more computational resources to aid in solving aerospace's more pressing challenges, new technologies like 3-d printing and additive manufacturing can be employed to manufacture new model pieces for small geometries like protuberances and fences in near real-time during large, expensive test campaigns.

Impact

There are a wide variety of areas where digitally transforming the scientific data supply chain will have a worthy impact: scientific research, facility and operations management, and engineering managers.

Scientific Research

Challenges foster a culture of open science (Brunton, et al., 2021). While companies and facilities may retain some algorithms, a simple step to be made is making input and output of files open to enable third-party software development. Connecting facilities and communities will facilitate advancements in model-based engineering. Demonstration of a digitally transformed scientific

data supply chain will initiate conversations in the ground testing community on how to increase the use of artificial intelligence to automate more of the scientific workflow and accelerate scientific discovery.

Facility and Operations Management

As a community, we must better characterize the tunnels because they lack the quality information on the acoustic environment is impeding the progress for programs accessing the facilities. Using high resolution optical technologies, we can characterize our facilities and learn information about our facilities that is currently unknown.

Given the ground testing facilities owned by the U.S. federal government are not-for-profit, the facilities are not responsible for ensure the best design for any given customer, however, they are responsible for designing, building, and offering a system that allows customers to explore more cost effectively. Some standard processing tools can be made open source, and enable competitive advantage to companies competing to win bids for NASA, Department of Defense, or similar.

Realization of Ground Testing 3.0 will take time, just like the development of CFD and technologies within the wind tunnel. It's a continuum.

Engineering Management

Many of the most impactful benefits of digitally transforming the scientific data supply chain were not revealed until the unparalleled disruptions of 2020 arrived. Indeed, 2020 was the year of disruptions. The COVID-19 pandemic took center stage globally. In addition, civil unrest in the spring of 2020 would motivate more Americans to join the Black Lives Matter Movement than had marched during the civil rights movement of the 1960s (Ruggs & Avery, 2021). Private and public companies and organizations responded by updating company policies and core values. NASA would add a fifth core value of inclusion to its list of safety, integrity, teamwork, and excellence. The way companies, including a U.S. federal agency like NASA, responded was manifested in a wide array of manners.

However, an idea that has not gained attention is how the new motivation behind digital transformation and a surge in the evolution of company diversity, equity, inclusion, and accessibility (DEIA) are very closely linked. Many times, when DEIA initiatives are discussed, recruitment and internships are the conversational focus; however, that only skims the surface of what DEIA should represent. By leveraging the benefits of a PESTEL analysis, we can amplify social DEIA initiatives by outlining what technical DEIA looks like in a digitally transformed

workplace, specifically the scientifically engineering workplace—technical diversity, technical equity, technical inclusion, and technical accessibility.

The value Project: Red Rover was amplified after the disruptions of 2020. The year 2020 highlighted that Project: Red Rover was a solution for a highly sociotechnical system. The facilities and the work are technical in nature, but they are driven by human actors.

“Digital transformation is more about the people than the technology” (Hill, Le Cam, Menon, & Tedards, 2022). Digital transformation will enable companies to better meet and expand their diversity, equity, inclusion, and accessibility (DEIA) initiatives. Today's engineering managers must be able to direct technical work, but equally important, they must continue to train themselves on how to manage the people under their supervision. The human actor in the technical organization is what makes the technical organization work and the reason for the organization's existence. Digital transformation includes the development of an organization's mindset and outlook that values skills like inclusiveness, openness, and teamwork (Deloitte, 2017).

Most of the time DEIA initiatives are discussed in a social context. However, just like addressing social issues, innovation can also address socio-technical challenges in the workplace. By innovating the transportation method of the scientific data supply chain, technical inclusion and technical accessibility can be enabled.

In addition, the post-2020 workforce has changed. Today's engineering managers should be aware and responsive to innovative technologies that allow their organizations to improve and adapt to the post-2020 workforce landscape.

- **Technical Inclusion:**

Ground testing facilities must embrace all of the core values (Bals & Tate, 2017) of their parent department or agency. NASA updated its core values in 2020 with the addition of inclusion. It will take time to understand and expand how inclusion is addressed with ground testing facilities. With the traditional scientific data supply chain, individuals depend on their network, and the information that an individual receives is dependent on their position within the network. It is very likely to have asymmetrical information channels because of this (Liao & Chen, 2017), which leaves some at a disadvantage compared to others on the quality and timeliness of information. Digitally transforming the scientific data supply chain enables technical inclusion by allowing better data management practices to give transparency to who has the data and when access is granted. This can be determined and transparent before the test campaign is run.

- **Technical Accessibility**

By removing the travel requirement to be present at the experimental facility to have a positive influence on the test campaign enables technical accessibility. With digital transformation, more individuals can have access to the data in real time and contribute positively to a test campaign regardless of their location. It has been noticed in practice that a group that is marginalized because of the requirement to travel is dependent care givers and ones with underlying health concerns.

- **Post-Pandemic Workforce**

The individuals entering the workforce have preconceived expectations that data is readily available and remotely accessible. The post-pandemic workforce has experienced a digitally transformed way to interact with colleagues irrespective of location. The scientific user facility has the control to digitally transform how data is delivered to customers.

- **Next-Generation Workforce**

The next generation of students are entering the workforce. They experienced a once-in-a-lifetime pandemic and saw more Americans actively protesting social unrest than had demonstrated during the 1960's civil rights movement (Ruggs & Avery, 2021). This generation has been greatly influenced and is deciding how to devote their talents and are genuinely devoted to making a positive impact in the world (Jónasson, Ramdas, & Sungu, 2022). Many desire to work for organizations that value their talents and value their benefit to society. They also understand there are many organizations to work for and will not settle for poor organizational values and managers.

The next-generation workforce also wants to excel in their work and continuously improve their skillsets. The well-known 10,000-hour rule goes much deeper than simply performing an action for 10,000 hours. The original research by Ericsson, Krampe, and Tesch-Römer (1993) titled "The Role of Deliberate Practice in the Acquisition of Expert Performance" brought to the mainstream by Gladwell (2008), documented that the 10,000 rule requires deliberate practice, with an instructor or mentor, giving immediate informative feedback, and knowledge of the results of their performance. Therefore violin, golf, and computer programming are often examples. The trainee in any of these subjects can immediately see the results of their performance. However, for a customer planning and executing a scientific test campaign with a post-test long lead time of two to four weeks, the deliberate practice with immediate feedback does not exist. Digitally transforming the scientific data

supply chain enables immediate feedback to the customer and in effect, better training for the next generation of scientists and engineers.

Engineering managers must realize everyone's contributions is unique but recognition it's not. recognition should not be solely based on papers or tools but rather on their contributions to the field. As designed and desired, individuals on cross disciplined teams will not produce the same output.

Currently, the workforce is experiencing a paradigm shift of working in the environment where the data resides, rather than pulling data to our local machine (Marx, 2013). Scientist and engineers must start to using enterprise level tools for computational purposes. That may take several years and some reverse mentoring.

Conclusion

To accelerate scientific discovery and increase knowledge creation during the design phase, the ground testing community must begin developing and deploying the technologies that will be required in 10-20 years. The future of aeroscience will continue to focus on the complex phenomena of unsteady aerodynamics. Therefore, the future of aerosciences will continue to increase efforts to develop, deploy, and adopt optical technologies delivering higher spatiotemporal resolutions required to study and record such complex phenomena over space and time. Given the increase in optical measurement and sophisticated sensors, the data volumes and velocities will continue to increase and the need and requirement for high-performance computing.

By viewing this series of events through the lens of supply chain management, we are able to apply well-studied models to improve the process of converting data to knowledge by improving the transportation of scientific data to enable information sharing among a large group of customer subject matter experts, great profits, that is knowledge created, is realized. Digitally transforming the scientific data supply chain has far reaching effect in scientific research at ground test facilities. It also enables facility managers to better maintain and service their facility, a national asset. Lastly, the biggest impact is transforming this system for the individuals who are using the system. Such a digitally transformed data supply chain enable technical accessibility and technical inclusion, allowing more individuals to participate and positively influence a ground test while the test is actively running in a facility while also enabling the next generation of engineers and scientist.

CHAPTER FOUR

CONCLUSION

As data volumes increase, so does the scientific research communities' aptitude to unlock scientific discovery. With the increase in data the scientific research community must ask itself, how do we harness the power of big data to increase knowledge creation and accelerate scientific discovery. The answer is revealed by viewing data through the lens of supply chain management and viewing data as the commodity that can be manufactured, distributed, and sold. Like any commodity it needs careful planning and coordination with all participants across the supply chain to ensure quality and timely delivery. By doing so, we establish the *scientific data supply chain*, which focuses on the management of scientific data. Many of the natural sciences are model rich and are well poised to exploit the large data sets by digitally transforming the scientific data supply chain.

The value of data will continue to increase in the field of scientific research. We have reached a new era in the ability to collect and analyze large datasets, revealing new discoveries across the natural sciences. The future of these discoveries will continue to focus on spatiotemporal phenomena, which will require data collection of large data sets in space and time. To do so, optical technologies have become a mainstay in this exploration and scientific discovery. While these images in their raw form can be fascinating to many, they are deemed rather useless to the scientific community without further data processing from raw form to a process data product. To add value to these invaluable data sets, the future of scientific research and discovery is high performance. Computational resources beyond our consumer-grade, personal machines are required. High performance computing resources are available to the scientific research community, but currently underused. Organizational structure must be built to connect experimental and computational communities and facilities to enable the increase in knowledge creation and accelerate scientific discovery.

The future of scientific research and discovery is cross-discipline. The Department of Energy's Advancing Scientific Knowledge Discovery Working Group Report (Department of Energy, 2013) highlights the importance of the multi-armed newsvendor model. The report states as hypotheses become more complex and cross-disciplinary, subject matter experts must be closely connected during scientific evaluation. The multi-armed newsvendor model addresses the need to make data available quickly and to a broad, cross-disciplinary audience, regardless of their location.

Lastly, given the cross-discipline audience, the future of scientific research and discovery will include participants will be a hybrid work force, some present at the facility when a test campaign is underway and many others remote. The fact that “Digital transformation is more about the people than the technology” (Hill, Le Cam, Menon, & Tedards, 2022), digitally transforming the scientific data supply chain is powerful and allows for social DEIA initiatives to be extended to the technical facets of the workplace. Technical accessibility removes the travel requirements to participate as a scientific user allowing more individuals to participate. Technical inclusion allows better data management practices and gives transparency to who has the data and when access was granted.

Beyond application of the multi-armed newsvendor model that emphasized allowance of shorter windows of data collection during a test campaign and enable real-time information sharing, future work will include automation more parts of the scientific workflow. As discussed, the data acquisition is highly automated and a contributing factor to the data deluge, but the scientific research community has reached the point where we should consider just because we can take data, doesn’t mean we should. Automating the data transfer from local storage to high-performance storage decreases the risk of data being stored on hardware that eventually becomes obsolete. Storing in a high-performance environment also allows more meta data to be stored and enables improved software engineering management practices. Automation of the data processing and employing data-driven optimization and data dimensionality reduction will push the scientific community towards real-time steering of experiments. This will require AI, but fortunately many fields of science are model rich, and now they are data rich. Brunton, et al. (2021) advocate advances in engineering and data science are driven by 1) large amounts of data, 2) advances in high-performance computing, 3) improvements in technology, including sensors, transfer and storage of data, 4) scalable, efficient algorithms, and 5) use and investment in open-source.

In conclusion, this research has motivated and defined the scientific data supply chain and delivered the multi-armed newsvendor model. The research has expanded the literature on the data supply chain originally motivated by Spanaki et al. (2018). By view scientific data as the commodity moving through the supply chain, we are able to realize the scientific research community’s goal with the increase of available data to increase knowledge creation and accelerate scientific discovery. The multi-armed newsvendor model has hybridized two well-known supply chain optimization models. With connecting the experimental and computational facilities, resources, and communities to increase knowledge creation and accelerate scientific discovery, our imagination is now becoming the limiting factor.

REFERENCES

- Abeykoon, V., Liu, Z., Kettimuthu, R., & Foster, I. (2019). Scientific image restoration anywhere. *2019 IEEE/ACM 1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP)*, 8-13. doi:<https://doi.org/10.1109/XLOOP49562.2019.00007>
- Arrow, K., Harris, T., & Marschak, J. (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, 19(3), 250-272. doi:<https://doi.org/10.2307/1906813>
- Aryal, A., Liao, Y., Nattuthurai, P., & Li, B. (2018). The emerging big data analytics and IoT in supply chain management: a systematic review. *Supply Chain Management*, 25(2), 141-156. doi:<https://doi.org/10.1108/SCM-03-2018-0149>
- Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47, 235-256. doi:<https://doi.org/10.1023/A:1013689704352>
- Baaziz, A., & Quoniam, L. (2015). How to use big data technologies to optimize operations in upstream petroleum industry. *International Journal of Innovation*, 1(1), 19-25. doi:<https://doi.org/10.5585/iji.v1i1.4>
- Bals, L., & Tate, W. (2017). Sustainable supply chain design (SSCD) in social businesses: advancing the theory of supply chain. *Journal of Business Logistics*, 39(1), 57-79. doi:<https://doi.org/10.1111/jbl.12172>
- Ban, G. Y., & Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1), 90-108. doi:<https://doi.org/10.1287/opre.2018.1757>
- Bansal, P., Gualandris, J., & Kim, N. (2020). Theorizing supply chains with qualitative big data and topic modeling. *Journal of Supply Chain Management*, 56(2), 7-18. doi:<https://doi.org/10.1111/jscm.12224>
- Bear, S., Parkinson, D., Gerhardt, L., Sibony, M., & Enders, B. (2021). ALS share, a lightweight data sharing service for a synchrotron radiation facility. *2021 IEEE / ACM 3rd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 1-7). doi:<https://doi.org/10.1109/XLOOP54565.2021.00006>
- Belyaev, N. L., Klimentov, A. A., Konoplich, R. V., Krasnopevtsev, D. V., Prokofiev, K. A., & Velikhov, V. E. (2017). High performance computing system in the framework of the Higgs boson studies at ATLAS. *XXVI International Symposium on Nuclear Electronics & Computing, 2023*, pp. 23-29. Becici, Budva, Montenegro. Retrieved from <https://ceur-ws.org/Vol-2023/23-29-paper-4.pdf>
- Benureau, F., & Oudeyer, P. -Y. (2016). Behavior diversity generation in autonomous exploration through reuse of past experience. *Frontiers in Robotics and AI*, 3(8), 1-24. doi:<https://doi.org/10.3389/frobt.2016.00008>
- Berenguer, G., Keskinocak, P., Shanthikumar, J., Swaminathan, J., & Van Wassenhove, L. (2017). A Prologue to the special issue on not-for-profit operations management.

- Production and Operations Management*, 26(6), 973-975.
doi:<https://doi.org/10.1111/poms.12730>
- Bicer, T., Nikitin, V., Aslan, S., Gürsoy, D., Kettimuthu, R., & Foster, I. (2020). Tomographic reconstruction of dynamic features with streaming sliding subsets. *2020 IEEE/ACM 2nd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 8-15). doi:<https://doi.org/10.1109/XLOOP51963.2020.00007>
- Bird, J., & Langelaan, J. (2020). A multi-armed bandit approach to atmospherically-aware altitude optimization. *AIAA Scitech 2020 Forum*. doi:<https://doi.org/10.2514/6.2020-1089>
- Biswas, R., Dunbar, J., Hardman, J., Bailey, F., Wheeler, L., & Rogers, S. (2012). The impact on high-end computing on NASA missions. *IT Professional*, 14(2), pp. 20-28.
doi:<https://doi.org/10.1109/MITP.2011.110>
- Bouneffouf, D., Rish, I., & Aggarwal, C. (2020). Survey on applications of multi-armed and contextual bandits. *2020 IEEE Congress on Evolutionary Computation (CEC)*, (pp. 1-8). doi:<https://doi.org/10.1109/CEC48606.2020.9185782>
- Bouneffouf, D., Rish, I., & Cecchi, G. (2017). Bandit models of human behavior: Reward processing in mental disorders. *International Conference on Artificial General Intelligence*. doi:https://doi.org/10.1007/978-3-319-63703-7_22
- Brinch, M. (2018). Understanding the value of big data in supply chain management and its business processes. 37(8), 1589-1614. doi:https://doi.org/10.1108_IJOPM-05-2017-0268
- Brunton, S., Kutz, N., Manohar, K., Aravkin, A., Morgansen, K., Klemisch, J., . . . McDonald, D. (2021). Data-driven aerospace engineering: Reframing the industry with machine learning. *AIAA Journal*, 8, 2820 - 2847. doi:<https://doi.org/10.2514/1.J060131>
- Brunton, S., Noack, B., & Koumoutsakos, P. (2020). Machine learning in fluid mechanics. *Annual Review of Fluid Mechanics*, 52, 477-508. doi:<https://doi.org/10.1146/annurev-fluid-010719-060214>
- Brunton, S., Proctor, J., & Kutz, N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences (PNAS)*, 113(15), 393.2-3937. doi:<https://doi.org/10.1073/pnas.1517384113>
- Byrne, J., & Miescher, D. R. (2019). The 150th anniversary of the discovery of DNA. *Biomics*, 11(3), 249-258. doi:<https://doi.org/10.31301/2221-6197.bmcs.2019-23>
- Canyakmaz, C., Özekici, S., & Karaesmen, F. (2022). A newsvendor problem with markup pricing in the presence of within-period price fluctuations. *European Journal of Operational Research*, 1, 153-162. doi:<https://doi.org/10.1016/j.ejor.2021.09.042>
- Carter, C., Rogers, D., & Choi, T. (2015). Toward the theory of the supply chain. *Journal of Supply Chain Management*, 51(2), 89-97. doi:<https://doi.org/10.1111/jscm.12073>

- Chakrabarti, D., Kumar, R., Radlinski, F., & Upfal, E. (2008). Mortal multi-armed bandits. (D. K. Bottou, Ed.) *Advances in Neural Information Processing Systems* 21, 273-280. Retrieved from <https://proceedings.neurips.cc/paper/2008/file/788d986905533aba051261497ecffcbb-Paper.pdf>
- Chatfield, D., Kim, J., Harrison, T., & Hayya, J. (2009). The bullwhip effect—impact of stochastic lead time, information quality, and information sharing: A simulation study. *Production and Operations Management*, 13(4), 340–353. doi:<https://doi.org/10.1111/j.1937-5956.2004.tb00222.x>
- Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, 32(4), 4-39. doi:<https://doi.org/10.1080/07421222.2015.1138364>
- Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, 46(3), 436-443. doi:<https://doi.org/10.1287/mnsc.46.3.436.12069>
- Chiu, C., & Choi, T.-M. (2016). Supply chain risk analysis with mean-variance models: A technical review. *Annals of Operations Research*, 240(2), 489-507. doi:<https://doi.org/10.1007/s10479-013-1386-4>
- Choppala, S., Kelmar, T., Chierichetti, M., Davoudi, F., & Huang, D. (2023). Optimal sensor location and stress prediction on a plate using machine learning. *AIAA SCITECH 2023 Forum*. National Harbor, Maryland. doi:<https://doi.org/10.2514/6.2023-0370>
- Clery, D., & Voss, D. (2005). All for one and one for all. *Science*, 809-809. doi:<https://doi.org/10.1126/science.308.5723.809>
- Coe, C. (1961). *Steady and fluctuating pressures at transonic speeds on two space-vehicle payload shapes*. NASA Technical Memorandum. doi:<https://ntrs.nasa.gov/api/citations/19650018342/downloads/19650018342.pdf>
- Cole, H. A., Erickson, A. L., & Rainey, A. G. (1970). *Buffeting during atmospheric ascent – space vehicle design criteria*. National Aeronautics and Space Administration.
- Cooke, N. J., & Hilton, M. L. (2015). *Enhancing the effectiveness of team science*. Washington, DC: National Academies Press (US).
- Coutant, A., Roper, K., Trejo-Banos, D., Bouthinon, D., Carpenter, M., Grzebyta, J., . . . King, R. D. (2019). Closed-loop cycles of experiment design, execution, and learning accelerate systems biology model development in yeast. *Proceedings of the National Academy of Sciences (PNAS)*, 116(36), 18142-18147. doi:<https://doi.org/10.1073/pnas.1900548116>
- Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *Proceedings of the 8th Conference on Visualization*, (pp. 235–244). doi:<https://doi.org/10.5555/266989.267068>

- Crispino, L. C., & Kennefick, D. J. (2019). A hundred years of the first experimental test of general relativity. *Nature Physics*, 15(5), 416-419. doi:<https://doi.org/10.1038/s41567-019-0519-3>
- Davenport, T. H., & Kirby, J. (2015). Beyond automation. Retrieved April 22, 2020, from <https://hbr.org/2015/06/beyond-automation>
- De Blas, J., Cepeda, M., D'Hondt, J., Ellis, R., Grojean, C., Heinemann, B., . . . Verkerke, W. (2020). Higgs boson studies at future particle colliders. *Journal of High Energy Physics*, 2020(1), 1-97. doi:[https://doi.org/10.1007/JHEP01\(2020\)139](https://doi.org/10.1007/JHEP01(2020)139)
- de Castro Moraes, T., & Yuan, X. M. (2021). Data-driven solutions for the newsvendor problem: A systematic literature review. *Advances in Production Management Systems* (pp. 149-158). Nantes, France: Springer International Publishing. doi:https://doi.org/10.1007/978-3-030-85910-7_16
- Deloitte. (2017). *Digital Future Readiness*. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/ch/Documents/consumer-business/ch-chip-en-swiss-transformation.pdf>
- Department of Energy. (2013). *Accelerating Scientific Knowledge Discovery (ASKD) Working Group Report*. doi:<https://doi.org/10.2172/1471107>
- Department of Energy. (2022). *Advanced Light Source Five-Year Strategic Plan FY2023*. Berkeley Labs, Office of Science. Retrieved from <https://als.lbl.gov/wp-content/uploads/2022/02/ALS-Strategic-Plan-FY2023.pdf>
- Dirks, D., & Stepro, N. (2023). Activating your healthcare data supply chain. *Healthcare Information and Management Systems Society (HIMSS)*. Retrieved from https://www.himss.org/sites/hde/files/handout-Ses-184_89.pdf
- Durand, A., Achilleos, C., Iacovides, D., Strati, K., Mitsis, G. D., & Pineau, J. (2018). Contextual bandits for adapting treatment in a mouse model of de novo carcinogenesis. *Proceedings of the 3rd Machine Learning for Healthcare Conference*, 85, pp. 67-82. Retrieved from <https://proceedings.mlr.press/v85/durand18a.html>
- Ebert, C., & Duarte, C. (2018). Digital transformation. *IEEE Software*, 35(4), 16-21. doi:<https://doi.org/10.1109/MS.2018.2801537>
- El Kadiri, S., Grabot, B., Thoben, K. D., Hribernik, K., Emmanouilidis, C., Von Cieminski, G., & Kiritsis, D. (2016). Current trends on ICT technologies for enterprise information systems. *Computers in Industry*, 79, 14-33. doi:<https://doi.org/10.1016/j.compind.2015.06.008>
- Elias, J. R., Chard, R., Libera, J. A., Foster, I., & Chaudhuri, S. (2020). The manufacturing data and machine learning platform: enabling real-time monitoring and control of scientific experiments via IoT. *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, (pp. 1-2). New Orleans, LA. doi:<https://doi.org/10.1109/WF-IoT48130.2020.9221078>

- Epstein, D. (2019). *Range: Why generalist triumph in a specialized world*. New York: Riverhead Books. Retrieved from <https://davidepstein.com/the-range>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406. doi:<https://doi.org/10.1037/0033-295X.100.3.363>
- Fathizadeh, F., Savinien, J., & Rekik, Y. (2022). Fuzzy Gaussian mixture optimisation of the newsvendor problem: Mixing fuzzy perception and randomness of customer demand. *International Journal of Production Research*, 1-22. doi:<https://doi.org/10.1080/00207543.2022.2085209>
- Fortunato, S., Bergstrom, C., Börner, K., Evans, J., Helbing, D., Milojević, S., . . . Vespignani, A. (2018). Science of science. *Science*, 359(6379). doi:<https://doi.org/10.1126/science.aao0185>
- Foxley, S., Sampathkumar, V., De Andrade, V., Trinkle, S., Sorokina, A., Norwood, K., . . . Kasthuri, N. (2021). Multi-modal imaging of a single mouse brain over five orders of magnitude of resolution. *NeuroImage*, 238, 118250. doi:<https://doi.org/10.1016/j.neuroimage.2021.118250>
- Frankel, R., Bolumole, Y. A., Eltantawy, R. A., Paulraj, A., & Gundlach, G. T. (2008). The domain and scope of SCM's foundational disciplines—Insights and issues to advance research. *Journal of Business Logistic*, 29(1), 1-30. doi:<https://doi.org/10.1002/j.2158-1592.2008.tb00066.x>
- Garbeff, T., Baerny, J., & Ross, J. (2019). Wind tunnel flow field visualizations of the space launch system vehicle ascent. *AIAA Aviation Forum*. doi:<https://doi.org/10.2514/6.2019-3299>
- Gentner, D., & Smith, L. (2012). Analogical reasoning. In *Encyclopedia of Human Behavior: Second Edition* (pp. 130-136). Elsevier Inc. doi:<https://doi.org/10.1016/B978-0-12-375000-6.00022-7>
- Gentner, D., Brem, S., Ferguson, R., Markman, A., Levidow, B., & Wolff, P. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of The Learning Sciences*, 6, 3-40. doi:https://doi.org/10.1207/s15327809jls0601_2
- Ghasemaghaei, M., & Calic, G. (2019). Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, 104, 69-84. doi:<https://doi.org/10.1016/j.jbusres.2019.07.006>
- Gittins, J., & Jones, D. (1979). A dynamic allocation index for the discounted multi-armed bandit problems. *Biometrika*, 66(3), 561-565. doi:<https://doi.org/10.2307/2335176>
- Gladwell, M. (2008). *Outliers: The story of success*. Little, Brown and Company. doi:<https://doi.org/10.1080/15210960903028818>

- Görgülü, B., & Sarhangian, V. (2022). A newsvendor approach to design of surgical preference cards. *Service Science*, 14(3), 213-230. doi:<https://doi.org/10.1287/serv.2021.0298>
- Gorzynski, J. E., Goenka, S., Shafin, K., Jensen, T., Fisk, D., Grove, M., . . . Kol. (2022). Ultrarapid nanopore genome sequencing in a critical care setting. *New England Journal of Medicine*, 386, 700-702. doi:<https://doi.org/10.1056/NEJMc2112090>
- Graves, S. C., Tomlin, B. T., & Willems, S. P. (2022). Supply chain challenges in the post-covid era. *Production and Operations Management*, 31(12), 4319– 4332. doi:<https://doi.org/10.1111/poms.13854>
- Gray, J. (2022). *The James Webb Space Telescope uses a 68GB SSD to store its incredible images and data*. Retrieved from DPReview.com: <https://www.dpreview.com/news/6078073136/the-james-webb-space-telescope-uses-a-68gb-ssd-to-store-its-incredible-images-data>
- He, T., Vicaire, P., Yan, T., Luo, L., Gu, L., Zhou, G., . . . Abdelzaher, T. (2006). Achieving real-time target tracking using wireless sensor networks. *12th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'06)* (pp. 37-48). IEEE. doi:<https://doi.org/10.1109/RTAS.2006.9>
- Hickner, M., Fasel, U., Nair, A., Brunton, B., & Brunton, S. (2023). Data-driven unsteady Aeroelastic Modeling for control. *AIAA Journal*, 61(2), 780-792. doi:<https://doi.org/10.2514/1.J061518>
- Hill, L., Le Cam, A., Menon, S., & Tedards, E. (2022). Digital transformation: A new roadmap for success. *Harvard Business School Working Knowledge*. Retrieved from <https://hbswk.hbs.edu/item/leading-in-the-digital-era-a-new-roadmap-for-success>
- Huber, J., Müller, S., Fleischmann, M., & Stuckenschmidt, H. (2019). A data-driven newsvendor problem: From data to decision. *European Journal of Operational Research*, 278(3), 904-915. doi:<https://doi.org/10.1016/j.ejor.2019.04.043>
- Islam, M., Roozeboom, N., Gunderson, P., Li, X., & Yu, A. (2022). An agent-based simulation model to mitigate the bullwhip effect via information sharing and risk pooling. *Winter Simulation Conference*. Singapore.
- Jónasson, J. O., Ramdas, K., & Sungu, A. (2022). Social impact operations at the global base of the pyramid. *Production and Operations Management*, 31(12), 4364–4378. doi:<https://doi.org/10.1111/poms.13857>
- Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. doi:<https://doi.org/10.1126/science.aaa8415>
- Kalkanci, B., Rahmani, M., & Toktay, L. (2019). The role of inclusive innovation in promoting social sustainability. *Production and Operations Management*, 28(12), 2960-2982. doi:<https://doi.org/10.1111/poms.13112>

- KC, D. S., Scholtes, S., & Terwiesch, C. (2019). Empirical research in healthcare operations: Past research, present understanding, and future opportunities. *Manufacturing & Service Operations Management*, 22(1), 73-83. doi:<https://doi.org/10.1287/msom.2019.0826>
- Kefalas, M., de Santiago Rojo, J., Apostolidis, A., van den Herik, D., van Stein, B., & Bäck, T. (2022). Explainable intelligence for exhaust gas temperature of turbofan engines. *Journal of Aerospace Information Systems*, 19(6), 447-454. doi:<https://doi.org/10.2514/1.1011058>
- Keller, R., Ranjan, S., Wei, M., & Eshow, M. (2016). Semantic representation and scale-up of integrated air traffic management data. *Proceedings of the International Workshop on Semantic Big Data* (pp. 1-6). Association for Computing Machinery. doi:<https://doi.org/10.1145/2928294.2928296>
- Kirenga, A., Lee, H., & Zolotas, A. (2023). Unmanned aerial system concept design for rail yard monitoring. *AIAA SciTech Forum*. National Harbor, MD. doi:<https://doi.org/10.2514/6.2023-1733>
- Kmak, F. (2000). Modernization and activation of the NASA Ames 11- by 11-Foot Transonic Wind Tunnel. *21st Aerodynamic Measurement Technology and Ground Testing Conference*. doi:<https://doi.org/10.2514/6.2000-2680>
- Korthäuer, T., Accorinti, A., Scharnowski, S., & Kähler, C. (2023). Effect of Mach number and pitching eigenfrequency on transonic buffet onset. *AIAA Journal*, 61(1), 112-124. doi:<https://doi.org/10.2514/1.J061915>
- Kuleshove, V., & Precup, D. (2000). Algorithms for multi-armed bandit problems. *Journal of Machine Learning Research*. doi:<https://doi.org/10.48550/arXiv.1402.6028>
- Lai, T. L. (1987). Adaptive treatment allocation and the multi-armed bandit problem. *Annals of Statistics*, 15, 1091-1114. doi:<https://doi.org/10.1214/AOS/1176350495>
- Lazarova-Molnar, S., Mohamed, N., & Al-Jaroodi, J. (2018). Collaborative data analytics for industry 4.0: Challenges, opportunities and models. *6th International Conference on Enterprise Systems (ES)* (pp. 100-107). IEEE. doi:<https://doi.org/10.1109/ES.2018.00023>
- Lee, S. (2022). Polymer research for scientific discovery and engineering applications. 47(1), 68.
- Levental, M., Chard, R., Libera, J., Chard, K., Koripelly, A., Elias, J., . . . Foster, I. (2020). Towards online steering of flame spray pyrolysis nanoparticle synthesis. *2020 IEEE/ACM 2nd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 35-40). doi:<https://doi.org/10.1109/XLOOP51963.2020.00011>
- Li, J., Lash, L., Roozeboom, N., Garbeff, T., Henze, C., Murakami, D., . . . Tang, L. (2022). Dynamic mode decomposition of unsteady pressure-sensitive paint measurements for the NASA Unitary Plan Wind Tunnel tests. *AIAA SciTech Forum*. doi:<https://doi.org/10.2514/6.2022-0141>
- Li, S., Chen, J., Liao, Y., & Shi, Y. (2016). The impact of information sharing and risk pooling on bullwhip effect avoiding in container shipping markets. *International Journal*

- Shipping and Transport Logistics*, 8(4), 406-424.
doi:<https://doi.org/10.1504/IJSTL.2016.077308>
- Liao, C. N., & Chen, Y. J. (2017). Farmers' Information Management in Developing Countries—A Highly Asymmetric Information Structure. *Production and Operations Management*, 26(6), 1207-1220. doi:<https://doi.org/10.1111/poms.12678>
- Lin, S., Chen, Y., Li, Y., & Shen, Z. J. (2022). Data-driven newsvendor problems regularized by a profit risk constraint. *Production and Operations Management*, 31(4), 1630-1644. doi:<https://doi.org/10.1111/poms.13635>
- Liu, Y., & Li, L. (2021). A map of bandits for e-commerce. *arXiv - CS - Machine Learning*. doi:[arxiv-2107.00680](https://arxiv.org/abs/2107.00680)
- Loiseau, J. C., & Brunton, S. L. (2018). Constrained sparse Galerkin regression. *Journal of Fluid Mechanics*, 838, 42-67. doi:<https://doi.org/10.1017/jfm.2017.823>
- Losada, D. E., Parapar, J., & Barreiro, A. (2017). Multi-armed bandits for adjudicating documents in pooling-based evaluation of information retrieval systems. *Information Processing & Management*, 53(5), 1005–1025. doi:<https://doi.org/10.1016/j.ipm.2017.04.005>
- Luthans, F., Luthans, B., & Luthans, K. (2015). *Organizational behavior: An evidence based approach* (13th ed.). IAP-Information Age Publishing.
- Marx, V. (2013). Biology: The big challenges of big data. *Nature*, 498(7453), 225-260. doi:<https://doi.org/10.1038/498255a>
- Matsuyama, K. (2006). The multi-period newsboy problem. *European Journal of Operational Research*, 171(1), 170-188. doi:<https://doi.org/10.1016/j.ejor.2004.08.030>
- Mendez, M., Ianiro, A., Noack, B., & Brunton, S. (2022). *Data-driven fluid mechanics: Combining first principles and machine learning*. Cambridge: Cambridge University Press.
- Micol, J. (2001). Langley Research Center's Unitary Plan Wind Tunnel: Testing capabilities and recent modernization activities. *39th AIAA Aerospace Sciences Meeting and Exhibit*. doi:<https://doi.org/10.2514/6.2009-1726>
- NASA. (2016). *Monometer Board Setup in Supersonic Wind Tunnel*. Retrieved from [nasa.gov: https://www.nasa.gov/image-feature/manometer-board-setup-in-supersonic-wind-tunnel](https://www.nasa.gov/image-feature/manometer-board-setup-in-supersonic-wind-tunnel)
- NASA. (2019). *From wind to data, in no time flat: accelerating spacecraft and aircraft design*. Retrieved Nov 6, 2019, from [nasa.gov: www.nasa.gov/image-feature/ames-from-wind-to-data-in-no-time-flat-accelerating-spacecraft-and-aircraft-design](https://www.nasa.gov/image-feature/ames-from-wind-to-data-in-no-time-flat-accelerating-spacecraft-and-aircraft-design)
- NASA. (2022, Sept). *Ames Unitary Plan Wind Tunnel*. Retrieved from [nasa.gov: https://www.nasa.gov/centers/ames/orgs/aeronautics/windtunnels/index.html](https://www.nasa.gov/centers/ames/orgs/aeronautics/windtunnels/index.html)

- NASA. (2022, July). *Pleiades Supercomputer*. Retrieved from nasa.gov: <https://www.nas.nasa.gov/hecc/resources/pleiades.html>
- Noack, M., & Zwart, P. (2019). Computational strategies to increase efficiency of Gaussian-process-driven autonomous experiments. *2019 IEEE/ACM 1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 1-7). doi:<https://doi.org/10.1109/XLOOP49562.2019.00006>
- Oak Ridge National Lab. (2022). *ORNL celebrates the launch of Frontier: the world's fastest supercomputer*. Retrieved Nov 7, 2022, from ornl.gov: <https://www.ornl.gov/news/ornl-celebrates-launch-frontier-worlds-fastest-supercomputer>
- Parkinson, D., Krishnan, H., Ushizima, D., Henderson, M., & Cholia, S. (2020). Interactive parallel workflows for synchrotron tomography. *2020 IEEE/ACM 2nd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 29-34). doi:<https://doi.org/10.1109/XLOOP51963.2020.00010>
- Peter, O., Pradhan, A., & Mbohwa, C. (2017). Industrial internet of things (IIoT): Opportunities, challenges, and requirements in manufacturing businesses in emerging economies. *Procedia Computer Science*, 217, 856-865. doi:<https://doi.org/10.1016/j.procs.2022.12.282>
- Petruzzi, N. C., & Dada, M. (1999). Pricing and the newsvendor problem: A review with extensions. *Operations Research*, 47(2), 183-194. doi:<https://doi.org/10.1287/opre.47.2.183>
- Piatak, D., Sekula, M., & Rausch, R. (2016). Sensitivity of Space Launch System buffet forcing functions to buffet mitigation options. *54th AIAA Aerospace Sciences Meeting*, (pp. 1-52). San Diego, California. doi:<https://doi.org/10.2514/6.2016-0544>
- Piatak, D., Sekula, M., Rausch, R., Florance, J., & Ivanco, T. (2015). Overview of the Space Launch System transonic buffet environment test program. *AIAA SciTech Forum*. doi:<https://doi.org/10.2514/6.2015-0557>
- Porteus, E. (1990). Foundations of stochastic inventory theory. In D. Heyman, & M. Sobel (Eds.), *Handbooks in Operations Research and Management Science* (Vol. 2, pp. 605-652). Amsterdam: Elsevier Science Publishers.
- Qiao, Y., Yang, W., & Sun, Z. (2022). A multi-product newsvendor problem model based on demand transfer and mental accounting in the context of e-commerce. *2022 3rd International Conference on E-commerce and Internet Technology (ECIT 2022)* (pp. 97-108). Atlantis Press. doi:https://doi.org/10.2991/978-94-6463-005-3_11
- Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research*, 213(2), 361-374. doi:<https://doi.org/10.1016/j.ejor.2010.11.024>

- Rafique, R., Mun, K., & Zhao, Y. (2017). Designing energy supply chains: Dynamic models for energy security and economic prosperity. *Production and Operations Management*, 26(6), 1120-1141. doi:<https://doi.org/10.1111/poms.12689>
- Rao, R. (2020, Sept 25). *Synchrotrons face a data deluge*. Retrieved from Physics Today: <https://physicstoday.scitation.org/doi/10.1063/PT.6.2.20200925a/full/>
- Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning*. Cambridge, Massachusetts: MIT Press. Retrieved from <https://gaussianprocess.org/gpml/>
- Rice University. (1962). *We choose to go to the moon*. Retrieved from Rice University: <https://www.rice.edu/jfk-speech>
- Richey, R., Morgan, T., Lindsey-Hall, K., & Adams, F. (2016). A global exploration of big data in the supply chain. *International Journal of Physical Distribution and Logistics Management*, 46(8), 710-739. doi:<https://10.1108/IJPDLM-05-2016-0134>
- Richins, G., Stapleton, A., Stratopoulos, T. C., & Wong, C. (2017). Big data analytics: Opportunity or threat for the accounting profession? *Journal of information systems*, 31(3), 63-79. doi:<https://doi.org/10.2308/isys-51805>
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin American Mathematical Society*, 58(5), 527-535.
- Rodgers, E., Hay, J., & Lussos, R. (2022). Coordinating innovative technology development at NASA. *ASCEND 2022*, (p. 4352). doi:<https://doi.org/10.2514/6.2022-4352>
- Romano, P. (2009). How can Fluid dynamics help supply chain manangement. *International Journal of Production Economics*, 118(2), 463-472. doi:<https://doi.org/10.1016/j.ijpe.2008.12.011>
- Roozeboom, N., & Baerny, J. (2017). Guide for pressure-sensitive paint testing at NASA Ames Research Center Unitary Plan Wind Tunnel. *AIAA SciTech 2017 Forum*. doi:<https://doi.org/10.2514/6.2017-1055>
- Roozeboom, N., Baerny, J. K., Murakami, D. D., Ngo, C., & Powell, J. M. (2020). Recent developments in NASA's unsteady pressure-sensitive paint capability. *AIAA Scitech 2020 Forum*. doi:<https://doi.org/10.2514/6.2020-0516>
- Roozeboom, N., Ngo, C., Powell, J. M., Baerny, J., Murakami, D., Ross, J. C., & Murman, S. (2018). Data processing methods for unsteady pressure-sensitive paint application. *AIAA SciTech Forum*. Kissimmee, Florida. doi:<https://doi.org/10.2514/6.2018-1031>
- Roth, A. M., & Singhal, V. R. (2022). Pioneering role of the Production and Operations Management in promoting empirical research in operations management. *Production and Operations Management*, 31(12), 4529– 4543. doi:<https://doi.org/10.1111/poms.13847>

- Roy, K., & Tomlin, C. (2008). Traffic flow management using supply chain and FIR filter methods. *AIAA Guidance, Navigation, and Control Conference and Exhibit*, (pp. 1-14). Honolulu, Hawaii. doi:<https://doi.org/10.2514/6.2008-7400>
- Ruggs, E., & Avery, D. (2021). Linking good intentions to intentional action. *MIT Sloan Management Review*, pp. 95-96. Retrieved from <https://sloanreview.mit.edu/article/linking-good-intentions-to-intentional-action/>
- Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., & Harnisch, M. (2015). Industry 4.0: The future of productivity and growth in manufacturing industries. *Boston Consulting Group*, 9(1), 54-89. Retrieved from https://www.bcg.com/publications/2015/engineered_products_project_business_industry_4_future_productivity_growth_manufacturing_industries
- Salim, M., Uram, T., Childers, J. T., Vishwanath, V., & Papka, M. (2019). Balsam: Near real-time experimental data analysis on supercomputers. *2019 IEEE/ACM 1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 26-31). Denver, CO. doi:<https://doi.org/10.1109/XLOOP49562.2019.00010>
- Schickore, J. (2014). Scientific discovery. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/Entries/scientific-discovery/>
- Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, 324(5923), 81-84. doi:<https://doi.org/10.1126/science.1165893>
- Schuster, D. M., & D'Agostino, M. G. (2017). *Future of NASA's aerosciences capabilities*. NASA Thermal and Fluids Analysis Workshop. Retrieved from <https://ntrs.nasa.gov/api/citations/20170009016/downloads/20170009016.pdf>
- Schuster, D., Panda, J., Ross, J., Roozeboom, N., Burnside, N., Ngo, C., . . . Piatak, D. (2016). *Investigation of unsteady Pressure-Sensitive Paint (uPSP) and dynamic loads balance to predict launch vehicle buffet environments*. NASA/TM-2016-219352. Retrieved from <https://ntrs.nasa.gov/api/citations/20160014849/downloads/20160014849.pdf>
- Schwartz, E. M., Bradlow, E. T., & Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4), 500-522. doi:<https://doi.org/10.1287/mksc.2016.1023>
- Sekula, M., Piatak, D., Rausch, R., Florance, J., & Ramey, J. (2015). Initial assessment of Space Launch System transonic unsteady pressure environment. *AIAA SciTech Forum*. doi:<https://doi.org/10.2514/6.2015-0558>
- Senning, J. (2020). CP343 History and overview of high-performance computing. Gordon College. Retrieved from https://www.math-cs.gordon.edu/courses/cps343/presentations/History_and_Overview_of_HPC.pdf
- Shen, W., Wang, J., Jiang, Y.-G., & Zha, H. (2015). Portfolio choices with orthogonal bandit learning. *24th International Conference on Artificial Intelligence*, (pp. 974-980). doi:<https://doi.org/10.5555/2832249.2832384>

- Shi, Y. (2022). *Advances in big data analytics*. Springer.
- Shih, T. I.-P. (2023). 2023 editorial—Doing more of the same is not enough. *AIAA Journal*, 61(1). doi:<https://doi.org/10.2514/1.J062678>
- Slivkins, A. (2019). *Introduction to multi-armed bandits*. Now Foundations and Trends. doi:<https://doi.org/10.48550/arXiv.1904.07272>
- Sotoudeh, Z., & Eger, M. (2023). Automatic adaptive refinement sampling method for data-driven predictive models. *AIAA SciTech 2023 Forum*. doi:<https://doi.org/10.2514/6.2023-0369>
- Sotoudeh, Z., Yuan, Z., & Barnes, C. (2023). Sparse identification of monlinear gust response in laminar flow. *AIAA SciTech Forum*. National Harbor, Maryland. doi:<https://doi.org/10.2514/6.2023-0367>
- Spanaki, K., Gürgüç, ., Z., Adams, R., & Mulligan, C. (2018). Data supply chain (DSC): Research synthesis and future directions. *International Journal of Production Research*, 56(13), 4447-4466. doi:<https://doi.org/10.1080/00207543.2017.1399222>
- Steva, T., Pollard, V., Herron, A., & Crosby, W. (2019). Space Launch System aeroacoustics wind tunnel test results. *AIAA Aviation Forum*, (pp. 1-19). Dallas, TX. doi:<https://doi.org/10.2514/6.2019-3303>
- Stremper, S., Yoshii, K., Hammer, M., Bycul, D., & Miceli, A. (2021). Designing a streaming data coalescing architecture for scientific detector ASICs with variable data velocity. *3rd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 8-14). doi:<http://doi.org/10.1109/XLOOP54565.2021.00007>
- Sun, Y., & Ma, O. (2023). Drone-based automated exterior inspection of an aircraft using reinforcement learning technique. *AIAA SciTech Forum*. National Harbor, Maryland. doi:<https://doi.org/10.2514/6.2023-0107>
- Sutton, R., & Barto, G. (2015). *Reinforcement learning: An introduction*. Cambridge, Massachusetts: M.I.T. Press. Retrieved from <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>
- Taira, K., Brunton, S., Dawson, S., Rowley, C., Colonius, T., McKeon, B., . . . Ukeiley, L. (2017). Modal analysis of fluid flows: An overview. *AIAA Journal*, 55(12), 4013-4041. doi:<https://doi.org/10.2514/1.J056060>
- Tereyağoglu, N., Fader, P., & Veeraraghavan, S. (2017). Pricing theater seats: The value of price commitment and monotone discounting. *Production and Operations Management*, 26(6), 1056-1075. doi:<https://doi.org/10.1111/poms.12611>
- Thayer, J., Damiani, D., Dubrovin, M., Ford, C., Kroeger, W., O'Grady, C., . . . Zohar, S. (2019). Data processing at the linac coherent light source. *2019 IEEE / 1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 32-37). doi:<https://doi.org/10.1109/XLOOP49562.2019.00011>

- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4), 285–294. doi:<https://doi.org/10.2307/2332286>
- Tonidandel, S., King, E. B., & Cortina, J. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods*, 21(3), 525–547. doi:<https://doi.org/10.1177/1094428116677299>
- TOP500. (2022). *TOP500 list – June 2022*. Retrieved November 7, 2022, from top500.com: <https://www.top500.org/lists/top500/list/2022/06>
- Tran-Thanh, L., Rogers, A., & Jennings, N. R. (2012). Long-term information collection with energy harvesting wireless sensors: A multi-armed bandit based approach. *Journal of Autonomous Agents and Multi-Agent Systems*, 25, 352–394. doi:<https://doi.org/10.1007/s10458-011-9179-0>
- U.S. Department of the Interior. (2017). *Unitary Plan Wind Tunnel*. Retrieved from National Park Service: <https://www.nps.gov/articles/unitary-plan-wind-tunnel.html>
- Vera C. Rubin Observatory. (2023). *About Rubin Observatory*. Retrieved from Legacy Survey of Space and Time (LSST): <https://www.lsst.org/about>
- Vescovi, R., Li, H., Kinnison, J., Keçeli, M., Salim, M., Kasthuri, N., . . . Ferrier, N. (2020). Toward an automated HPC pipeline for processing large scale electron microscopy data. *2020 IEEE/ACM 2nd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)*, (pp. 16–22). doi:<https://doi.org/10.1109/XLOOP51963.2020.00008>
- Villari, B. C., Abdulla, M. S., Sebastian, M. P., & Ramaprasath, L. (2020). *Novel multi-armed bandit algorithm for business applications*. Indian Institute of Management Kozhikode. Retrieved from <http://dspace.iimk.ac.in/xmlui/handle/2259/1059>
- Waymo. (2023). *Waymo Driver*. Retrieved from Waymo: <https://waymo.com/>
- Wei, C.-C., & Chen, L.-T. (2021). Supply chain replenishment decision for newsvendor products with multiple periods and a short life cycle. *Sustainability*, 13(22), 12777. doi:<https://doi.org/10.3390/su132212777>
- Wen, Z., Kveton, B., Valko, M., & Vaswani, S. (2017). Online influence maximization under independent cascade model with semi-bandit feedback. *Advances in Neural Information Processing Systems*, 3022–3032. doi:<https://doi.org/10.5555/3294996.3295062>
- Wieland, A., & Durach, C. F. (2021). Two perspectives on supply chain resilience. *Journal of Business Logistics*, 42(3), 315–322. doi:<https://doi.org/10.1111/jbl.12271>
- Winski, C., Danehy, P., Watkins, A. N., Shea, P., Meerooff, J., Lowe, K. T., & Houlden, H. (2019). Space Launch System booster separation supersonic powered testing with surface and off-body measurements. *AIAA Aviation Forum*. Dallas, Texas: AIAA. doi:<https://doi.org/10.2514/6.2019-3505>

- Wozniak, J., & Schwarz, N. (2019). Message from the workshop chairs. *2019 IEEE/ACM 1st Annual Workshop on Extreme-Scale Experiment-in-the-Loop Computing (XLOOP)*, (p. 5). doi:<https://doi.org/10.1109/XLOOP49562.2019.00004>
- Xiao, L., & Wang, C. (2023). Multi-location newsvendor problem with random yield: Centralization versus decentralization. *Omega*, *116*, 102795. doi:<https://doi.org/10.1016/j.omega.2022.102795>
- Yang, C., Hu, Z., & Zhou, S. X. (2021). Multilocation newsvendor problem: Centralization and inventory pooling. *Management science*, *67*(1), 185-200. doi:<https://doi.org/10.1287/mnsc.2019.3532>
- Zhang, J., & Bareinboim, E. (2017). Transfer learning in multi-armed bandits: A causal approach. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, (pp. 1340-1346). doi:<https://doi.org/10.24963/ijcai.2017/186>

VITA

Nettie H. Roozeboom is a practicing engineer and project manager in the aerospace industry. Upon graduation from the University of Tennessee at Chattanooga with her undergraduate in mechanical engineering, she began her career at NASA Ames Research Center in Moffett Field, CA in 2008. She completed her master's degree part-time at Stanford University in mechanical engineering in 2011. She conducted air-traffic management research and worked operations in the NASA Ames Unitary Plan Wind Tunnel. In 2012, she began her research in the optical measurements field with an emphasis on Pressure-Sensitive Paint (PSP). She led the deployment of the PSP technology on production tests at the NASA Ames Unitary Plan Wind Tunnel. In 2014, technology would enable the ability to collect PSP data at high speeds (20 kHz). At that point, Roozeboom began demonstrating the capability of the technology on the Space Launch System (SLS) vehicle. Since 2019, she has been the principal investigator leading the unsteady PSP measurement technology and Project: Red Rover, a project connecting world-class experimental and computational facilities at NASA Ames Research Center. This work inspired the work documented here.