LEVERAGING BREAKTHROUGHS IN DATA SCIENCE TO OPTIMIZE READINESS DECISION-MAKING

LEVERAGING BREAKTHROUGHS IN DATA SCIENCE TO OPTIMIZE READINESS DECISION-MAKING

THE CHALLENGE: Overcoming the Limits of Traditional Approaches to Readiness Assessment

With peer and near-peer adversaries closing in on the United States military's technological and numerical advantages and their increasingly more aggressive posturing, there is little room for error in today's readiness planning and resource allocation. Using available data, including both qualitative and quantitative information, military commanders today draw heavily upon their knowledge and experience to develop an intuitive understanding of the factors that go into allocating resources to achieve desired readiness levels in areas such as training, personnel and equipment. But no single person, or even a group of decision-makers, can fully consider the complex interrelationships of the hundreds, perhaps thousands of variables that may underlie a readiness investment decision.

Increasingly, military commanders and their civilian counterparts are expected to tease out the complex tradeoffs in terms of both costs and benefits to optimize the Department of Defense's (DoD) readiness. However, allocating resources (e.g., personnel, equipment and training) to meet objective, or individually measured, readiness

metrics can lead to suboptimal results in understanding aggregate readiness. In part, this is due to a limited ability to assess the marginal readiness achieved from resource decisions (dollars spent) as well as the potential second- and third-order effects of a given activity (not to mention consideration of the expected operational environment and threat).

In addition to assessing the operational impacts of current readiness conditions, commanders are also expected to constantly and efficiently balance the available resources to maintain and improve readiness to meet near- and longer-term requirements. The thinking is, a commander who can better understand the costs and expected readiness gains associated with various training events, in a quantitative and comparative model, will be more informed and better able to make and defend difficult tradeoff decisions. This kind of fact-based assessment can not only lead to more effective use of resources, but also help justify budgets and make the case for funding.

The factors used by military leaders today to determine readiness include qualitative and quantitative assessments ranging from commanders' reports to measures of personnel availability, training, equipment usage, maintenance cycles, supply chains and even human factors. Readiness reporting systems across the DoD collect

current readiness inputs and near-term projections from military units with the goal of generating objective, accurate and timely qualitative and quantitative readiness measures.

While these systems contain a profusion of readiness-related data, they are struggling to provide decision makers responsible for assessing composite readiness with a sufficient understanding of the complex interrelationships among readiness variables or the tools for determining where to best spend the next "readiness dollar." This is, in part, due to the data and analytic infrastructures used by most defense organizations today. With these infrastructures, readiness data tends to be locked in isolated silos that are hard to connect and integrate. One of the key problems is that the infrastructures were designed before the age of big data. They simply weren't intended to handle the volume, velocity and variety of the readiness data collected today.

Just preparing the data for analysis tends to be a cumbersome, labor-intensive process. Surveys have found that IT professionals spend as much as 80-90 percent of their time in data preparation – such as collecting, cleaning, and formatting – leaving only 10-20 percent of their time for actual analysis.[1] This drain on resources makes it too costly and time-consuming for organizations to bring together all of their available readiness data – they're typically confined to narrow slices of data and insights.

Conventional infrastructures have prevented defense organizations from taking full advantage of predictive analytics, artificial intelligence (AI) and other data science approaches. While these groundbreaking technologies have the ability to find hidden patterns and interconnections, drill deeply into cause and effect and even predict the likely impact of a particular readiness decision, many organizations find themselves limited to just the first two phases in the evolution of analytics – descriptive and lower-level diagnostic.

Descriptive analytics sort through and summarize raw data to make it understandable, through spreadsheets, charts, reports and other kinds of presentations – showing current levels of manpower, training or equipment, for example. These analytics are also used to examine past trends, with the aim of guiding future actions.

While descriptive analytics can show what has happened in the past, or what is happening now, they don't explain *why* things happen. That's the role of diagnostic analytics, which look for interrelationships and connections. However, because most defense organizations are not able to fully integrate their full range of datasets, their diagnostic analytics can play only a limited role. For example, a diagnostic analytic might reveal the impact of one aspect of training on another. But since training data is typically separate from personnel data or equipment data, for example, the analytic can't get at the larger readiness picture – it can't easily cross those boundaries.

The defense community has made major strides with the first two phases of analytics. There is now an opportunity for many organizations to transition to the higher levels of analytics – ones that have the power to give decision-makers the insights and traceability they need to optimize their readiness planning and resource allocation.

A NEW PERSPECTIVE: Bringing Data Science to the Mission Organically and Dynamically

As defense organizations have sought to develop objective, data-driven readiness, they have often taken the approach of collecting ever-greater amounts and different kinds of data. But the problem isn't a lack of data – for the most part, defense organizations are already collecting all the data they need to fuel predictive and other powerful analytics. The challenge comes in bringing the data together as a whole, understanding how it is interrelated and uncovering deeper insight and foresight.

Advances in data science are now making it possible for defense organizations to accomplish this. With new approaches to infrastructure – such as open data platforms and data lakes – organizations can now quickly scale and integrate all their available data. This includes both structured and unstructured data from across functions and domains, and from any number of outside sources. And with advanced analytics, organizations can see how all this data comes together to create the big picture, and how that picture might change with different decisions.

Leveraging Data Science Infrastructures

One of the key advances in data science is that isolated data silos – and extensive data preparation – are no longer necessary. Open data platforms and other innovative infrastructure approaches now make it possible to break down the silos and bring together the entirety of the data for analysis. Instead of going through the slow, cumbersome process of building and rebuilding limited use relational databases, organizations can quickly tag and ingest a virtually unlimited amount of data into a highly secure, reusable data lake. This includes not only structured data, but unstructured and

semi-structured data as well. With data science, what formerly took days, weeks, or months to process can now be done in hours, minutes, seconds – or even microseconds. With this speed and scale, organizations can do exponentially more analysis at the same cost. Just as important, analysts are freed up to spend their time finding high-value insights, rather than laboriously formatting data.

The Power of Advanced Analytics

Such advances create enormous opportunities for readiness planning. Lower-level diagnostic analytics – which look at only limited datasets – cannot easily show the full set of interrelationships in a given situation. They often cannot answer the key questions commanders are asking, such as, "What are the hidden causes and effects at play here? What is driving the outcomes I'm seeing?" But because high-level diagnostics can access an organization's entire realm of data, these questions can now be answered.

With high-level diagnostics, readiness decision-makers, operators and analysts can quickly model and visualize the complex interrelation of information now isolated in data silos. For example, an analytic might show how 10 factors related to personnel and 10 factors related to training work in various combinations to increase or decrease readiness in a particular unit under varying conditions.

With the next level of analytic evolution – predictive – we can adjust inputs to see how the picture we've created in the diagnostic is likely to change if we take a particular action. Predictive analytics go far beyond the common practice of manually extrapolating from spreadsheets. These advanced analytics rely on computer models to create and run through any number of possible scenarios, and assign each one a likelihood of occurrence based on data about how events have unfolded in the past.

This kind of foresight can help organizations tease out complex readiness tradeoffs – in near real-time. Predictive analytics can reveal the impact of specific investments or cuts or show commanders how they can save money without impacting readiness. Or, a commander might use the analytic in the reverse way – plugging in the desired outcome, and then asking how the various contributing factors would have to be brought together. In this case, the commander might ask, "What are all of the ways I can cut costs without decreasing readiness?"

It's important to note that predictive and other advanced analytics are not intended to replace leadership decision-making. Instead, analytics are decision support tools that give commanders more information to work with. The experience, knowledge and judgment commanders bring is still essential to evaluate the analytic outputs and make final decisions. Today, commanders may spend much of their time trying to understand the relationships between isolated silos of information, such as data on equipment, training and maintenance. With advanced analytics, those relationships are already clear – freeing up commanders to focus on higher-level decisions.

Starting with People, Not Technology
What makes these and other data science capabilities possible is incorporating, at a granular level, the nature of the individual organization and its unique readiness mission requirements. Starting with the organization's people – including decision-makers, operators and analysts – is critical in bringing together their knowledge, experience and insight to help design and implement a data science-based approach. It builds trust in the environment because it reflects the needs and nuances of their organization and mission and is not a generic solution "customized" after the fact.

OUR APPROACH: Building in the Needs of the Organization and Mission from the Ground Up

Booz Allen Hamilton believes that defense organizations can take full advantage of data science for readiness planning and resource allocation, through a combination of infrastructure and analytic approaches that are technologically feasible, highly cost-effective and practical to adopt. This is now true even for organizations currently without the resources to stand up their own data science infrastructures, thanks to new shared services models. By bringing together industry-leading expertise in both data science and government technology consulting, Booz Allen brings a proven, step-by-step methodology for operationalizing data science in readiness planning.

Engaging Users
One of our first steps is to bring together and engage the organization's three key readiness stakeholder communities – the decision-makers, operators and analysts. Each group typically has its own priorities for data science. Senior leaders are looking to enhance readiness decision support with a broader, more holistic view of the enterprise that can get their questions answered quickly. Operators, such as planners and schedulers, want greater oversight as they track and monitor daily activities across the enterprise. And readiness analysts are looking to data science to expand their toolboxes in a broad range of capabilities, such as in modeling and simulation.

Booz Allen helps incorporate the stakeholder communities and their priorities into the process as a unified whole. We achieve this through our multi-disciplinary data science teams, which combine expertise in domain, computer science and mathematics. Our domain experts have a deep understanding of both the organization and its

mission. To bring data science to readiness, we call upon our experts in all of the key areas – personnel, equipment, training, supply chains, etc. – along with Booz Allen staff with experience working with the individual organization. We also bring highly skilled computer scientists to build the underlying computing environment, and AI-savvy mathematicians to create the analytics that will meet the organization's needs.

Assessing the Current State

Working with the stakeholder communities, we assess the organization's data and analytic capabilities in readiness planning. For example, what kinds of data are being collected, and how is it currently classified, stored and accessed? Who are the owners of the data, and what concerns do they have about sharing it? At the same time, we also look at where organizations are in the evolution of analytics. For example, what kinds of readiness insights are organizations getting from their analytics now? Are the analytics showing connections between the data stovepipes, or do readiness decision-makers, operators and analysts have to infer those connections on their own?

Envisioning the End State

In the next step, we help the readiness communities determine what hope to gain from data science. It's not enough to get together some data and throw a predictive analytic at it – that's only likely to yield random insight that may not be particularly helpful to the mission. In order to create data science infrastructures and analytics that build in the needs of the organization and its mission, you have to be clear at the outset where you're going and why. What kinds of readiness insights do you want? How do you want to use them? What kinds of dashboards and other visualizations will work best for your organization?

Through this process, known as envisioning, our teams help create a picture of the desired end state and definition of success. And because we speak the language of the enterprise, we can show the stakeholder communities the art of the possible.

Roadmapping

We then build a detailed project plan, or roadmap, for meeting target outcomes, encompassing all technical and non-technical aspects. The roadmap includes a project schedule, key milestones and time-phased deliverables, and is based on agile best practices, which help organizations build their data science capabilities through constant feedback and frequent iteration. The agile approach often calls for building and testing prototypes, which can then be expanded upon and refined as data science takes on a greater role in readiness planning. An agile mindset not only enables organizations to continually improve their data science, it helps achieve quick wins, which build engagement and buy-in.

Implementation

In the implementation phase, Booz Allen operationalizes the key aspects of data science for readiness planning, including infrastructure, data ingest and curation, analytics and visualizations.

Infrastructure

Booz Allen has developed several ways that defense organizations can now quickly, securely and cost-effectively access the infrastructures that underlie the data science environment. Whether organizations prefer to have sole access to their infrastructures or choose to take advantage of a shared services model, in which they securely bring their own data and analytics to a common infrastructure, we enable organizations to securely scale their data and employ plug-and-play modular data and analytic solutions using open data platforms. With this open approach, organizations are freed from vendor-lock, long-term licensing agreements, and expensive, proprietary solutions. Open data platforms allow organizations to quickly and economically switch capabilities in and out as

conditions, requirements and technologies change. In addition, organizations do not need to replace their legacy data systems – all of their available data can quickly be ingested into a data lake. Open data platforms are typically based in secure commercial cloud networks, which are significantly cheaper, and offer defense organizations substantially more data science opportunities, compared with traditional data centers.

Booz Allen works with organizations to operationalize their preferred best path to a data science infrastructure. For organizations that want sole access, we quickly set up a cloud-based infrastructure, in line with DoD's risk management framework (RMF) security controls. From there, we help the organization to achieve compliance, and obtain the necessary accreditations based on best practices we've developed over more than a decade of bringing advanced data science to the defense and intelligence communities.

For organizations who choose to stand up their own infrastructures we pursue a shared services model. With this approach, Booz Allen develops a common infrastructure that can be securely used by any number of partnering organizations. Since the environment already has an authority to operate (ATO) and all the necessary security controls, defense organizations can very quickly and inexpensively begin harnessing their data with AI and other advanced analytics.

Data Ingest and Curation

Once the infrastructure has been established, the next step is data ingest. Through our domain knowledge and working closely with the organization's stakeholders, we help organizations identify the datasets to be brought into the data lake. With data science infrastructures, defense organizations no longer need to cull their data when asking questions – all of the data can be used, and all at once. What's more, they can even go beyond their traditional data sources. That is, in addition to

securely bringing in readiness-related data from across the DoD, open data platforms can tap into a limitless variety of external, publicly available sources, ranging from weather to financial to social media.

Working collaboratively with the organization, Booz Allen also ensures that as the data is ingested into the data lake, it's tagged in a way that makes it discoverable and understandable. We work to ensure, for example, that readiness data from five different systems, describing a capability in five different ways – and even with five different spellings – can be tagged in a way that recognizes what the data means.

Analytics and Visualizations

Domain knowledge and collaboration is also critical in selecting the right datasets to train the readiness analytics. If analytic models are trained with inaccurate data, the outputs are likely to be faulty. Often, key datasets needed for training the analytics simply don't exist, requiring data scientists to build them from scratch – using synthesized data to fill in gaps where data is non-standard or of low quality.

Our collaboration with the decision-makers, operators and analysts continues in development of visualization tools, which enable each of the stakeholder communities to drill deep into the data for actionable readiness insight. One of the breakthroughs of data science is that people without specialized computer expertise can ask questions in ordinary language, without needing to know specialized programming languages. Booz Allen has developed tools that make this possible with the ease of an everyday Internet search. We work with the stakeholder communities to select the right tools for their specific needs, and to closely tailor those tools to the individual readiness mission. We also provide organizations with the ability to design their own "self-service" analytics – without the help of data scientists – to create new, ad hoc dashboards or other visualizations as needed.

DEFENSE ORGANIZATIONS CAN TAKE FULL ADVANTAGE OF DATA SCIENCE FOR READINESS PLANNING AND RESOURCE ALLOCATION THROUGH A COMBINATION OF INFRASTRUCTURE AND ANALYTIC APPROACHES

BOOZ ALLEN: Your Essential Partner In Operationalizing Data Science for Readiness

Booz Allen is an industry pioneer in bringing data science to government. We worked with the intelligence community – the earliest government adopter of advanced analytics – and are now helping a broad range of defense and other government organizations make a smooth, cost-effective transition to data science infrastructures and analytics. Our success is rooted in our ability to bring together a combination of deep domain and mission understanding, top-tier technical expertise and a culture of innovation.

As a readiness innovation leader, our staff have focused on innovative improvements, technical systems and operational enhancements. Booz Allen's Readiness Center of Excellence – Mission Readiness Capability focuses on formalizing best practices and applying innovation to deliver readiness solutions with increased speed and schedule efficiency.

In addition, we manage a broad portfolio of government contract vehicles to provide fast, flexible channels to access the full range of our technology and management services – including data optimization. Operating with an enterprise mindset – bringing together technical experts in analytics and data science, with operational experts with readiness experience – we can help you find the right vehicle for your needs. Together we can develop the right combination of infrastructure and analytic approaches that are technologically feasible, highly cost effective and practical for your organization to adopt.

OUR EXPERTS

For more information please contact our experts:

Sandeep Dorawala
Vice President
Dorawala_Sandeep@bah.com

Graham Evans
Principal
Evans_Graham@bah.com

Dave Merrill
Lead Associate
Merrill_Dave@bah.com

NOTES

1. Xplenty survey: *https://www.xplenty.com/news/2015-07-01-bi-professionals-spen-50-90-percent-of-their-time-cleaning-raw-data-for-analytics*
CrowdFlower/Figure Eight survey: *https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf*

About Booz Allen

For more than 100 years, business, government, and military leaders have turned to Booz Allen Hamilton to solve their most complex problems. They trust us to bring together the right minds: those who devote themselves to the challenge at hand, who speak with relentless candor, and who act with courage and character. They expect original solutions where there are no roadmaps. They rely on us because they know that — together — we will find the answers and change the world. To learn more, visit BoozAllen.com.