Data Science Teams:
Evolution of the Full-Contact Sport

by Colleen McCue, Ph.D. and Lt Col Brandon Daigle

No one today questions whether data science can confer competitive advantage to organizations effectively adopting these capabilities. Considerable attention has been paid to technology, talent, and training as organizations attempt to build data science programs in an effort to convert aspiration to practice; moving from simply counting and reporting what happened to data-informed prediction and prescription. As understanding of the various and specific data science roles has evolved, talent development, management, and leadership are emerging as new frontiers. Organizations now work to create data science capacity where effective and responsible use of advanced analytics makes the whole truly more than a sum of the parts.

Organizations increasingly recognize data science as a "must have" capability and are scrambling to recruit and hire data science talent. Data literacy programs are also proliferating given growing empirical research demonstrating the importance of upskilling across the organization in order to optimize data science talent and efforts. At the same time, the community is moving away from the data science "unicorn" concept in favor of a data science ecosystem. This is built through aggregation and assembly of complementary data science skill sets in support of meaningful, scalable, and sustainable breadth and depth of capacity.

The challenge many organizations now face is how to organize and manage this disparate yet complementary talent? While the data science Venn Diagram has become ubiquitous, recent analysis suggests we are actually looking at something with far greater complexity and depth– more akin to a data science "tapestry" where unique data science capabilities and roles are intricately interwoven across the enterprise.[297] Although aesthetically appealing, the tapestry construct is inherently static. Experience increasingly suggests that fixed data science resource allocation constructs are inflexible and even brittle; breaking in response to rapid evolution or change. Moreover, fixed or otherwise inflexible organizational constructs may have limited capacity to adjust on the fly to address unanticipated challenges and opportunity alike. What is needed, therefore, is data science capacity that is relatively fluid, agile and dynamic. It should be able to rapidly pivot in response to changing conditions, including unforeseen shortcomings or failures, while also setting the conditions to fully capitalize on unexpected windfalls in support of true competitive advantage. With this evolving understanding of high-performing data teams, we propose that analytically competitive organizations actually operationalize data science as something closer to a full-contact team sport where no one rides the bench.

THE FULL-CONTACT TEAM SPORT

Although others have described data science as a "team" sport, our experience shows that analytically competitive organizations extend the metaphor by establishing

"Efficiency remains important, but the ability to adapt to complexity and continual change has become an imperative."[298]

– Team of Teams

data science as a "full-contact" team sport where no one rides the bench.[299] Full-contact refers not to the physical domain, but rather the creation of a culture and environment where ideas are openly challenged and refined, and unflinching peer review is the norm.

In our experience, full intellectual "contact" is particularly important – it relates to identifying and mitigating cognitive bias, challenging assumptions and inference, and responsible interpretation and use of results. Based on this construct, "contact" should occur early and often, and involve all interested players. Tools such as the Heilmeier Questions can be invaluable for early evaluation, vetting and shaping.[300] Such tools establish the structure and discipline necessary for successful, actionable innovation, and getting the best ideas across the finish line.[301] On the other hand, unflinching and independent peer review should be ubiquitous given its value. Rigorous peer review in particular brings at least three separate, yet complementary benefits to the team.[302] First, it enables each individual performer to put their best work forward. Confirmation bias and error blindness are almost impossible to find in one's own work. Unflinching, independent, "full-contact" peer review can be used to identify unsupported assumptions, faulty logic, and errors in inference. While constructive feedback might sting, taking a few hits during practice can be invaluable if it sets the conditions for peak performance under game conditions. Culture should therefore be nurtured and modulated in order to find the sweet spot between soul-crushing criticism and overly congenial feedback that does not strengthen performance of the team.[303] Second, by providing peer review, the reviewer actively engages their higher cognitive functions giving themselves the type of mental workout necessary to improve their own performance. Just as every day is a "leg day" at the gym, every day should be a "brain day" for knowledge workers; providing critical peer review to a colleague affords an excellent opportunity to work the muscle between our ears. Finally, a program of rigorous peer review elevates both the overall quality of work and the "brand" of the work unit or team. Done consistently, this increases the overall quality and value of work products produced, which will translate into future opportunities.

ENTERPRISE-WIDE DATA SCIENCE: NO ONE RIDES THE BENCH

The idea that no one rides the bench refers to the importance of enterprise-wide upskilling in data literacy. Described as "the ability to make good judgments, use tools responsibly and effectively, and ultimately make good decisions using data," data acumen in particular is emerging as a "must-have" skill given its importance to responsible and effective use of data.[304] This idea is not new and extends from research supporting the importance of analytic maturity or readiness as an essential component of data science initiatives.[305] More

recently, organizations are beginning to realize the importance of data literacy and fluency, particularly as it relates to providing broader lift to enterprise-wide efforts that approach data literacy as a team sport rather than an isolated service center.[306] In addition to establishing data literacy as a means by which to achieve real competitive advantage, the data science community also appreciates the larger value proposition, which includes the potential monetary return on investment associated with upskilling the work force and developing analytic maturity.[307] This is not meant to suggest that everyone should learn to write [insert your favorite/trending language here] code. Rather, it simply means every member of the team should have a foundation in data literacy that enables them to participate in a meaningful, role-specific manner. Whether it is using their domain expertise to identify and shape requirements, contribute to testing and evaluation, or identifying potential algorithmic bias, data science is a team sport, and everyone plays.

LEADERSHIP: THE COACH AND MANAGERS

Filling leadership roles poses a perplexing challenge similar to that of achieving team-wide data literacy. While you would not expect the offensive line coach to run onto the field and take a snap, we routinely require data science leaders to retain currency in each new programming language, technology or software capability. Polly Mitchell-Guthrie, a data science thought leader who has created and led high performing data science teams, notes that hiring managers often mistakenly look for a long list of very specific technical skills when filling data science leadership roles.[308] While no one denies that hands-on, direct experience is necessary in data science leader roles, proficiency in every novel or otherwise intriguing capability becomes less important when compared with the ability to read the field, call plays, elevate and lead a team, and deliver results. Experience and perspective become even more important within the context of multiple, interrelated teams or working groups. Moreover, many successful coaches know the difficulty of trying to lead a team as a functioning member. Effective management of the team therefore requires some degree of separation, particularly when executing against organizational strategy that supersedes the vision and goals of a smaller working group.[309] At the end of the day, while a data science leader sitting in the corner knocking out code with the rest of the team might seem like a good idea, they likely will not have the separation and perspective necessary to shepherd the team across the finish line in support of the long game. This represents a poor use of the leadership role.

On the other hand, research underscores the importance of upskilling executive leadership roles in data literacy as a means to set conditions for data science success and manage expectations.[310] Candid discussion of what advanced analytics can and cannot do and the understanding that analytic capacity development is an incremental process can help manage expectations. This creates the necessary time and space to establish meaningful, scalable, and sustainable analytic capacity, especially given the proliferation of "easy button" solutions and the desire to sprinkle "AI/ML" like pixie dust across the enterprise.[311]

RULES OF THE GAME AND THE PLAYBOOK

While each role may play a different position, they all need to work from a common playbook informed by foundation-level data literacy and a shared understanding of individual roles and contributions to the game. A source, method, and technology-agnostic analytic workflow akin to a scientific method or checklist best supports this approach.[312] In practice, these data science "rules of the game" can set the conditions for the intellectual agility required to evolve in response to changing conditions, including seamlessly incorporating novel capabilities and those over the current horizon. This positions the team to effectively "read the field," identify opportunity, and run the gaps. Many have noted that innovation does not necessarily involve the creation of new science or math. Rather, it requires the vision to effectively capitalize on the transdisciplinary nature of data science; identifying novel uses for existing capabilities or repurposing successful approaches from one domain to another in support of novel insight and meaningful solutions to hard problems. Finally, this insight allows the team to seamlessly recover from broken plays and fumbles, or nimbly capitalize on an opponent's mistake because they have the requisite data literacy and insight to think on their feet and call out opportunity. Tools like the Heilmeier Questions can provide the type of structure, guidance, and soft restraint necessary to ensure that projects move down the field, advance, and are completed without soul-crushing restrictions.[313] Underscoring the importance of shared understanding and a common organizing playbook, we created posters and deskside reference cards with our workflow and process. We then distributed these to analysts and leaders alike as a means to both remind and emphasize the importance of reproducible analytic process and workflow.[314]

PUT ME IN, COACH!

In our experience, data science talent needs to be in the game, neither on the sidelines nor sequestered in the booth, remotely calling plays. This does not necessarily mean that data science experts take every snap, or even play on game day. Rather, direct access to end users and workflow provides the insight required for operationally-relevant and actionable analysis and associated model tuning, as well as the ability to see how or even if developed solutions are being implemented. It also affords opportunities for small, instructional "show me how" proof of concept or pilot studies.[315] Finally, one incredibly important benefit of embedding data science talent in the team is that it greatly increases access to this resource. This includes opportunities to identify, reinforce, and replicate organically emerging best practices, as well as affording early intervention and mitigation of challenges. In total, this sets the conditions for spontaneous "drive-by" engagement with the data scientist. Commercial best practice supports this model, indicating that sequestering data science talent in ivory towers or isolated service centers limits the ability to fully realize the promise of this resource.[316]

Innovation does not necessarily involve the creation of new science or math. Rather, it requires the vision to effectively capitalize on the transdisciplinary nature of data science.

TRAINING

The need for repetition, education, and training cannot be overstated. This is particularly true for the critical thinking and problem-solving skills that are so essential to the effective and responsible use of advanced analytics.[317] Again, identifying parallels to physical training can be illuminating. For example, no one would complete a great workout and believe they are ready for the season, yet many organizations continue to send their teams to "one and done" training despite a body of knowledge indicating these skills are perishable. Ongoing education, challenge, and practice are necessary to ensure that the brain stays sharp and in good condition. Moreover, growing experience indicates that learning by doing represents a great way to both internalize and creatively operationalize learning.[318] While time in the team room going over plays on the chalk board might be important, at some point the team needs to hit the turf and operationalize its learning. Whether education and training are executed through role-specific individualized learning journeys, regular group training, independent study, or making an organizational commitment to rigorous peer review, the benefits associated with creating and maintaining a culture of learning are without limit.

The special missions community in particular understands that teams train for the known but must educate for the unknown. Extending this analysis, we have written previously about the importance of developing the ability for teams to maintain cognitive performance under stress or pressure, which requires advance preparation and an ongoing commitment to education, training, and development. This often includes overtraining to unconscious competence or "flow"; a phenomenon frequently associated with elite athletes, but also known to improve cognitive performance.[319] Again, this is not a "one and done" activity. Whether a quarterly sales deadline or high-risk mission where lives hang in the balance, constant reps areessential to ensuring peak analytic performance under pressure.

THE IMPORTANCE OF "FAIR" PLAY: AI ETHICS AND RESPONSIBLE USE

In addition to knowing the rules of the game, general understanding of "fair" play is necessary. Just as winners never cheat and cheaters never win, biased algorithms simply do not work. While some rules are obvious—"I won't knowingly cheat"—others are more subtle and can be sport-specific. "Fair play" in data science can be even more challenging given the opaque nature of some algorithms, as well as the role that inherently flawed human cognition plays in assumption, judgement, inference, and heuristics across and throughout the process. Adding to the challenge is the conventional, albeit incorrect wisdom that advanced analytics to include AI and machine learning are inherently "objective" and can be used to eliminate bias. As former Principal Deputy Director of National Intelligence (PDDNI) Sue Gordon has noted, however, algorithms are simply "someone's opinion written in code."[320] Therefore, education in ethics and responsible use, or the concept of "fair play," becomes even more imperative. Further incentivizing this point, ethics or "fair play" in data science is emerging as a key differentiator in analytically competitive organizations that realize the promise of advanced analytics as compared to those simply chasing opportunity.[321] Again, education in critical thinking is key and organizations, including the U.S. Department of Defense are explicitly calling for education in the identification, understanding, and mitigation of AI/ML bias.[322] Limitations and constraints on the use of advanced analytics in the future will ensure meaningful,

effective, and responsible use of these powerful capabilities.

CONCLUSION

As organizations pivot from so-called data science "unicorns" to data science ecosystems, increasing evidence supports the important role of organization-wide data literacy upskilling.[323] Crucially, these ecosystems incorporate an array of disparate, yet complementary knowledge, skills, and experience in support of novel, transdisciplinary approaches to hard problems. Therefore, the question is not *if*, but *how* to upskill and organize talent. Growing evidence suggests that organization and talent management matters… a lot. The concept of a data science tapestry where professionals are intricately interwoven or affixed like precious jewels may be intuitively appealing. In our experience, however, effective and responsible data science is best practiced as a full contact sport where no one rides the bench. Teams can benefit further from this analogy by drawing lessons related to the role of leaders as a coach rather than player, the value of achieving shared understanding of the rules of the game through an organizational playbook, the importance of position players and training, and the concept of "fair play". Ultimately, this model speaks to the need for agile data science capacity that can rapidly and effectively evolve and respond to changes now and tomorrow.

[297]Fleming, O., Fountaine, T., Henke, N and T. Saleh (2018). Ten red flags signaling your analytics program will fail. McKinsey & Company. 4MAY18, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail
[298]McChrystal, S., Collins, T., Silverman, D. and Fussell, Cl (2015). Team of teams: New rules of engagement for a complex world. New York: Penguin Random House.
[299]Groeneveld, K. (2019). Why the data science Venn Diagram is wrong. LinkedIn, https://www.linkedin.com/pulse/why-data-science-venn-diagram-wrong-kees-groeneveld/?trackingId=%2FGEMgQBhXfsrgxNsshVD8w%3D%3D
[300]The Heilmeier Catechism. Defense Advanced Research Projects Agency (DARPA).
[301]Pisano, G.P. (2019). The hard truth about innovation cultures. Harvard Business Review. JAN-FEB19. https://hbr.org/2019/01/the-hard-truth-about-innovative-cultures
[302]McCue, C. and Daigle, B.J. (2019). Tip of the spear thinking: Maximizing analytical performance under pressure. Trajectory Magazine, 14JUN19, https://trajectorymagazine.com/tip-of--the-spear-thinking/
[303] Pisano, G.P. (2019). The hard truth about innovation cultures. Harvard Business Review. JAN-FEB19. https://hbr.org/2019/01/the-hard-truth-about-innovative-cultures
[304]National Academies of Sciences, Engineering, and Medicine (2018). Data science for undergraduates: Opportunities and Options. Washington, DC: The National Academies Press. https://doi.org/10.17226/25104.
[305]Davenport, T.H. and Harris, J.G. (2007). Competing on analytics: The new science of winning. Boston: Harvard Business School Publishing.
[306]Data Leader's Guide to Upskilling (datacamp). https://www.datacamp.com/resources/whitepapers/data-leaders-guide-to-upskilling
[307]Data Leader's Guide to Upskilling (datacamp). https://www.datacamp.com/resources/whitepapers/data-leaders-guide-to-upskilling
[308]Mitchell-Guthrie, P. (2021). Who is fit to lead data science? KDnuggets. 09FEB21. https://www.kdnuggets.com/2021/02/fit-lead-data-science.html
[309]Pisano, G.P. (2019). The hard truth about innovation cultures. Harvard Business Review. JAN-FEB19. https://hbr.org/2019/01/the-hard-truth-about-innovative-cultures
[310]Fleming, O., Fountaine, T., Henke, N and T. Saleh (2018). Ten red flags signaling your analytics program will fail. McKinsey & Company. 4MAY18, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail
[311]Davenport, T.H. and Jarvenpaa, S.L. (2008). Strategic use of analytics in government. IBM Center for the Business of Government. http://www.businessofgovernment.org/sites/default/files/Strategic%20Analytics.pdf
[312]McCue, C. (2014). Data mining and predictive analysis: Intelligence gathering and crime analysis (2nd ed). Waltham, MA: Butterworth-Heinemann.
[313]The Heilmeier Catechism. Defense Advanced Research Projects Agency (DARPA). https://www.darpa.mil/work-with-us/heilmeier-catechism
[314]McCue, C. and Daigle, B.J. (2019). Tip of the spear thinking: Maximizing analytical performance under pressure. Trajectory Magazine, 14JUN19, https://trajectorymagazine.com/tip-of-the-spear-thinking/
[315]Davenport, T.H. and Harris, J.G. (2007). Competing on analytics: The new science of winning. Boston: Harvard Business School Publishing.
[316]Data Leader's Guide to Upskilling (datacamp). https://www.datacamp.com/resources/whitepapers/data-leaders-guide-to-upskilling
[317]McCue, C. and Daigle, B.J. (2019). Tip of the spear thinking: Maximizing analytical performance under pressure. Trajectory Magazine, 14JUN19, https://trajectorymagazine.com/tip-of-the-spear-thinking/
[318]Boser, U. (2020). Learning by doing: what you need to know. The Learning Curve. 09MAR20. https://www.the-learning-agency-lab.com/the-learning-curve/learning-by-doing
[319]Kerr, J. (2013). Legacy: What the All Blacks Can Teach Us about the Business of Life. London: Constable.; and Kahneman, D. (2011). Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.
[320]As quoted in, McCue, C. and Daigle, B.J. (2019). Tip of the spear thinking: Maximizing analytical performance under pressure. Trajectory Magazine, 14JUN19, https://trajectorymagazine.com/tip-of-the-spear-thinking/ ; and, Kaplan, M.D.G. (2018). Building to scale and "creating anew". Trajectory Magazine, 24APR18, https://trajectorymagazine.com/building-to-scale-and-creating-anew/
[321]AI Momentum, Maturity & Models for Success (SAS, Accenture Applied Intelligence, and Intel with Forbes Insights). https://www.accenture.com/t20180914T183027Z__w__/us-en/_acnmedia/PDF-86/Accenture-AI-Momentum.pdf
[322]2020 Department of Defense Artificial Intelligence Education Strategy, Delivered in Fulfillment of the Requirements outlined in Section 256 of the National Defense Authorization Act for Fiscal Year 2020 (Public Law 116-92). SEP20, https://www.ai.mil/docs/2020_DoD_AI_Training_and_Education_Strategy_and_Infographic_10_27_20.pdf
[323]Data Leader's Guide to Upskilling (datacamp). https://www.datacamp.com/resources/whitepapers/data-leaders-guide-to-upskilling

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government.

Dr. Colleen McCue is a Principal Data Scientist with CACI International, where she supports special missions. She earned her doctorate in psychology from Dartmouth College, and completed a five-year postdoctoral fellowship at the Medical College of Virginia, Virginia Commonwealth University.

Lt Col Brandon Daigle is an Air Force Officer currently serving as a National Defense Fellow to the Fletcher School of Law and Diplomacy, Tufts University. He has held a variety of in-garrison and deployed leadership positions. He most recently served as the Commander of a Joint Unit at Ft Bragg, North Carolina. He holds a M.S. in Defense Analysis/Special Operations and Irregular Warfare, Naval Postgraduate School, a M.S. in Organizational Leadership and Design from Amridge University and a B.S. in Religion from Southern Christian University.