

Harvard Data Science Review • Issue 1.2, Fall 2019

Immersive Collaboration on Data Science for Intelligence Analysis

Alyson Wilson¹ Matthew Schmidt¹ Lara Schmidt² Brent Winter³

¹Laboratory for Analytic Sciences, North Carolina State University, Raleigh, North Carolina, United States of America,

²Strategic and Global Awareness Directorate (Project West Wing), Aerospace Corporation, El Segundo, California, United States of America,

³University Communications and Marketing, North Carolina State University, Raleigh, North Carolina, United States of America

The MIT Press

Published on: Nov 30, 2020

DOI: <https://doi.org/10.1162/99608f92.4a9eef8d>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

In 2013, the National Security Agency (NSA) founded the Laboratory for Analytic Sciences (LAS) at North Carolina State University (NCSU) to help the Intelligence Community (IC) address the growing complexity of big data challenges. The goal of LAS is to partner experts and practitioners from academia, government, and industry to create tools and techniques that help intelligence analysts provide better information to the decision makers who need it. This interview of Alyson Wilson, the principal investigator of LAS from NCSU, and Matthew Schmidt, the LAS technical director from NCSU, is conducted by Lara Schmidt, the principal director for Strategic and Global Awareness at the Aerospace Corporation, and Brent Winter from University Relations at NCSU. It provides an overview of the LAS collaboration model and describes projects conducted at LAS.

Keywords: anticipatory thinking, collaboration, innovation, intelligence community, structured analytic techniques

Brent Winter (BW): Alyson, could you tell us about what you are trying to accomplish at LAS?

Alyson Wilson (AW): We're focused on intelligence analysis. Broadly speaking, the main purpose of intelligence analysis is to provide information to policymakers that can help illuminate their decision options (Johnson, 2010; [ODNI, 2019](#)). At LAS, we're trying to improve intelligence analysis by breaking down some of the barriers between the IC and innovators at universities and companies. You know the stereotype from movies and TV shows—government intelligence agencies housed in fortress-like hidden facilities, where analysts with top-secret clearances analyze information gathered from halfway around the world? Well, that image isn't entirely false, but the IC has realized that every company has the same challenge they do, which is to figure out how to use data to gain strategic advantage; so they decided to partner with academia and industry to see if they could accelerate innovation in data science.

BW: So, how does this work? Is all your work classified, never to see the light of day?

AW: No, not at all. Even though LAS is supported by NSA and other IC partners, fully 90% of the lab's work is unclassified. NSA has a visible presence on NC State's campus. LAS and IC leadership believe that an open exchange of ideas will be able to overcome barriers to innovation and the integration of leading-edge academic and industry research into the IC.

Matthew Schmidt (MS): The government director of LAS, Michael Bender, is also a visiting research scholar and former associate faculty member at Loyola University. LAS has about 40 government staff, including NSA intelligence analysts and researchers, working on site. The government staff work in the lab on staggered 3-year rotations to help develop tools and techniques that would be useful to them in their regular intelligence

duties. These are no-kidding intelligence professionals; it is a real treat for academics to get to work closely with them to build collaborations to solve real-world problems.

Lara Schmidt (LS): So, let's talk a bit about the data science piece. You say you're trying to better inform decision makers through solid analysis. Can you give us a sense of what kinds of decisions you're talking about and how data science comes into play?

MS: One really interesting line of work that we had was a project associated with the 2016 summer Olympics in Rio de Janeiro. In the run-up to the Olympics, Brazil was experiencing a great amount of social unrest, with multiple marches and riots occurring daily throughout the country. The idea behind our work was to enhance the security of Olympics attendees by using publicly available information to predict the locations of large-scale public demonstrations in the Rio area. LAS researchers wrote software, called FareWays, to continuously acquire data from three sources: Rio's public transit system, which tracks buses' real-time locations via GPS sensors; a new municipal light-rail system built specifically for the Olympics; and publicly available traffic alerts for the Rio area. FareWays combined this data into a live stream viewed by system operators. It also included software that used topological data analysis for feature extraction to continuously analyze incoming data to predict trouble spots. Our analytic results were displayed as an overlay on top of a map of the Rio area in a secure operations center at the Olympics, staffed by intelligence personnel from all over the world. We proved that it could predict the locations of riots and other social disruptions, but as a rapidly developed prototype, the system wasn't quite ready for use in a live scenario. Still, it was an exciting opportunity for the LAS team. We think approaches like this could prove useful to public authorities all over the world as they grapple with the disruptions caused by societal unrest and riots. For instance, the 2011 London riots caused an estimated £200 million worth of property damage and resulted in five deaths and more than 3,000 arrests. Predicting such demonstrations in advance could help authorities reduce or prevent property destruction, injuries, and loss of life.

AW: The Rio Olympics example is just one of our projects related to anticipatory thinking. Basically, we're trying to re-envision risk analysis from the ground up. Risk is typically defined as the impact of an event times the likelihood of the event occurring. With our anticipatory thinking work, we are trying to develop better ways to envision an event's impact and avoid pitfalls such as failure of imagination. For instance, the [9/11 Commission Report](#) noted that before the 9/11 attacks, the North American Aerospace Defense Command (NORAD) had already performed an antiterrorism exercise that imagined attackers using airplanes as weapons—but the exercise had the planes departing from foreign cities (National Commission on Terrorist Attacks upon the United States, 2004). The failure to imagine that attackers might use *domestic* flights was a handicap that impaired the country's ability to defend against such a threat. Viewing this from a data science standpoint, there are two metrics often used to define classification accuracy: precision, a measure of how many identified results are relevant; and recall, a measure of the proportion of relevant results correctly identified. Forecasting focuses on precision, but anticipatory thinking focuses on recall. The difference between the two approaches is

one of mindset. Forecasting tries to lead the analyst to the most likely outcome, but anticipatory thinking uses an open, imaginative approach to envision all possible outcomes, not just the most likely ones.

Our initial interest was in determining whether anticipatory thinking was a capability that could be scientifically described, measured, and eventually taught to analysts. We started from cognitive science to derive and validate the constructs used to imagine possible futures and then built on these constructs to create metrics and training.

LS: That’s really interesting, and it sounds like some of the kinds of structured thinking approaches that intelligence professionals typically use. Are you leveraging data science to innovate in this area?

AW: Yes, exactly. For instance, the IC makes frequent use of structured analytic techniques (SATs), which are defined analytical methodologies intended to counteract humans’ known cognitive and perceptual limitations. At LAS we’re using data science to automate these SATs to increase their speed and, hopefully, their relevance to decision makers. One particularly promising tool is called the BEAST (Broadening and Enlightening Analytic Structured Tradecraft). The idea behind the BEAST is to use data science to help analysts bring more rigor to their decisions. Analysts are trained in a variety of defined rigorous analytical approaches, but in the real world they often wind up using whatever approach seems most feasible in the limited time available, sometimes at the loss of rigor. The BEAST team set out to develop a computerized tool that would incorporate some of that formal rigor and pass it along to the analyst in the form of practical, rapid decision support. The SAT upon which the BEAST is based is called ‘analysis of competing hypotheses,’ which is designed to help people evaluate multiple hypotheses that *could* explain a given set of observed data. The BEAST uses data science—specifically natural language processing, entity extraction, term frequency-inverse document frequency (tf-idf) similarity scoring, and optimized search and retrieval—to identify and push potential values to the analyst.

MS: The analyst uses the BEAST by entering a question they want to answer, in natural language. For example: *What is the chance the candidate xx will win the upcoming election in country yy?* The analyst can also supply more context, enter any relevant assumptions, and enter hypothetical answers. Based on those entries, the system retrieves as much data as it can, processes it through a relevance engine, and displays the results to the analyst, who then evaluates the data and decides whether it supports or refutes their various hypotheses. The analyst can delete refuted hypotheses, enter new ones, and provide additional assumptions, all of which becomes the context for the next data retrieval. The analyst iterates with the system in an iterative loop of hypotheses, data retrieval, and analysis until the analyst feels they have reached a high-quality decision with adequate rigor in the time available.

BW: So, is the BEAST aiming to make intelligence analysts obsolete? Are they going to go the way of the tax preparer in the post-TurboTax world?

AW: Importantly, the tool does *not* try to automate the human out of the loop. Instead, it tries to arm the human with more information so they can do what they're best at: critical thinking.

MS: The BEAST's data-triage model replaces the traditional intelligence tradecraft of querying data sets based on search terms. That traditional approach had the downside of reinforcing and amplifying any of the analyst's unconscious biases that might be embedded in the queries themselves. Instead, the BEAST uses data science to find *data* that may be relevant to the question of interest and then forces the analyst to create a hypothesis to explain what they're seeing in the data. This *broadening* of the field of inquiry (the B in BEAST) counteracts the natural human tendency to narrow in the face of limited time and competing cognitive demands. The BEAST seeks to help the analyst iterate between broadening and narrowing until a robust convergence is reached and a solid answer is developed.

BW: Have you tried this out in real-world settings? And, if so, what kind of results do you get?

MS: BEAST is still a proof-of-concept prototype, but LAS has conducted workshops where government analysts have used it to perform mission-relevant tasks to evaluate how it affects their work. After using the BEAST's iterative broadening-narrowing process, many analysts realized their work was based on assumptions they didn't know they had. Once they entered those assumptions into the tool, it returned data that refuted them, which led the analysts to increase the rigor of their analysis by tamping down some of those hidden biases that were affecting their work.

AW: We're also looking at other problems where data can be combined with, for example, social science research. We have a team working to better understand the risk factors for radicalization and terrorism. They began with a literature review of all published research on these risk factors, and they soon reached the disheartening conclusion that there wasn't much empirical evidence to review. One of the problems with developing that kind of evidence is that all data on radical terrorists comes from case studies of people who have already been identified as such. For example, we use data from the Western Jihadism Project, which is a database built from publicly available sources on known terrorism-involved individuals who spent formative years in the West ([Klausen, 2019](#); [McFee, Jensen, & James, 2019](#)). These data allow you to describe your terrorist population, but if you don't have data on comparable peers who haven't been radicalized, you can't do predictive analysis to figure out who among a non-radicalized population might go down the radical path. That's why the team is now using a variety of data sets—some publicly available, others available through data-sharing agreements—to link data on terrorists with matched samples of counterparts who haven't been radicalized. The goal is for researchers to ultimately identify evidence-based pathways to radicalization.

LS: It sounds like you are having some early successes by developing data-driven tools to augment what is already in intelligence analysts' toolkits. But I'm sure there were interesting bumps in the road along the way, right?

AW: We do think we've made a lot of progress since 2013 when LAS was founded, and we continue to add industry and academic partners from across the United States and even internationally. But, yes, there were definitely some interesting bumps in the road. It turns out that intelligence analysts, academic faculty, and private-industry professionals each take completely different approaches to their work. LAS's faculty partners tended to prefer working on small projects with their students; the analysts came from a culture of secrecy, so they didn't say much in meetings; and private industry was accustomed to waiting for the client to simply write a requirement for what they need. It was hard to get people collaborating the way we wanted. After a while, we came to the realization that we should be leveraging faculty partners with expertise in communications and organizational behavior to help us build a more effective collaboration.

MS: We called these new partners the Collaboration Team. While they initially limited themselves to observation and facilitation, they soon identified structural problems they thought they could help fix. They spent a couple of years performing interviews and observational studies with LAS project teams, collecting data across a wide range of teams and projects. Eventually they discovered reliably better ways for project teams to perform standard tasks like structuring a team, kicking off a project, running communications, and shutting projects down.

They made recommendations to the leadership about communicating the expectations around interdisciplinary work and creating incentives, structured opportunities, and rules of engagement to make it happen. They helped teams construct common, mutually beneficial, and intellectually interesting goals to rally around. The teams have implemented those improvements to great effect, leading many in the lab to conclude that the Collaboration Team's findings could be useful in settings well beyond LAS and the IC. The Collaboration Team has a forthcoming book that collects all the scholarship they've created in the course of helping LAS optimize its operations (Jameson, Tyler, Vogel, & Joines, in press).

AW: Over time, we have built a three-way collaboration model between the IC, industry, and academia, including about 45 grad and undergrad students. Students say what they value most about the experience is that they're working on a real problem in the same room with a real IC analyst, an academic expert, and a private-industry innovator, all of whom are facing the same problem. We call this our 'immersive collaboration' model. This kind of hands-on experience helps students develop a concrete understanding of what they're doing and why it's important, providing a foundation they can build upon to become data science leaders who will be in ever-increasing demand.

BW: So, what's next for LAS?

MS: We've talked about our work in anticipatory thinking. But we're also making progress along five other research thrusts: human-machine collaboration, advancing analytic rigor, machine learning and artificial intelligence, data triage, and cybersecurity. The common theme here is leveraging data science to make the intelligence analyst more effective. We're identifying computational tasks that would be beneficial to an

analyst, performing those tasks, and seamlessly integrating the results into the analyst's workflow. We're identifying ways to automatically characterize signals of interest harvested from sensor-enabled devices connected to the internet; at the same time, we're trying to strengthen policies and technologies that protect the privacy of sensitive information shared across these devices. No matter what question is of interest to decision makers and intelligence analysts, we're trying to ensure that data science increases the rigor and speed of intelligence analysis, saving analysts' time for tasks only a trained human can perform.

AW: Now that we are a year into our next 5-year contract, the lab is continuing to broaden its academic and industry collaborations beyond the initial focus on the Research Triangle area. LAS's distinctive three-way collaboration model is still the order of the day: government analysts, academic researchers, and private-industry innovators rubbing shoulders on project-focused teams, developing better ways to reap, parse, and analyze an ever-growing body of data to help decision makers address the security concerns facing us all.

LS: Thank you so much for speaking with us. It's been fun learning about one of the ways the IC is innovating in data science and the LAS immersive collaboration model.

Disclosure Statement

Alyson Wilson, Matthew Schmidt, Lara Schmidt, and Brent Winter have no financial or non-financial disclosures to share for this interview.

References

Jameson, J. K., Tyler, B. B., Vogel, K. M., & Joines, S. (in press). *Facilitating interdisciplinary collaboration among the intelligence community, academy, and industry*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Johnson, Loch K. (2010). National security intelligence. In L. K. Johnson (Ed.), *The Oxford handbook of national security intelligence* (pp. 3–32). Oxford, UK: Oxford University Press.

Klausen, J. (2019). *The Western Jihadism Project: An archive data file charting the evolution of Al Qaeda-inspired terrorist networks and recruitment in Western states, 1990 to the present*. Waltham, MA: Brandeis University. Retrieved September 17, 2019, from <https://www.brandeis.edu/klausen-jihadism/about.html>

McFee, G., Jensen, M., & James, P. (2019). *Profiles of individual radicalization in the United States (PIRUS)*. College Park, MD: National Consortium for Terrorism and Responses to Terrorism, University of Maryland. Retrieved September 17, 2019, from <https://www.start.umd.edu/data-tools/profiles-individual-radicalization-united-states-pirus>

National Commission on Terrorist Attacks upon the United States. (2004). *The 9/11 Commission report: Final report of the National Commission on Terrorist Attacks upon the United States*. Authorized ed., 1st ed. New York: Norton.

Office of the Director of National Intelligence (ODNI). (2019). *What is intelligence?* Retrieved from <https://www.dni.gov/index.php/what-we-do/what-is-intelligence>.

©2019 Brent Winter, Lara Schmidt, Matthew Schmidt, and Alyson Wilson. This interview is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the interview.