

Data Science Center Tames Big Data Projects

GREG NEGUS
CORNERSTONE RESEARCH

► **Greg Negus is the chief operating officer at Cornerstone Research. In this interview, he shares innovations in big data, as well as how artificial intelligence and machine learning can be used to support expert testimony.**

Can you tell us a little about yourself and your role at Cornerstone Research?

Greg Negus: I have enjoyed a thirty-year career in professional services firm management. Prior to joining Cornerstone Research, I worked as a chief operating officer and chief financial officer at several large law firms.

At Cornerstone Research, I am responsible for leading the firm's corporate and administrative functions. I also serve on Cornerstone Research's executive committee, among others. In this position, I also oversee the firm's Data Science Center, which is an interdisciplinary team of in-house data scientists. The Data Science Center team's expertise includes improving work efficiencies and implementing state-of-the-art modeling as it relates to analysis of relevant data and case issues.

Can you talk about the goals of Cornerstone Research's Data Science Center?

At the firm level, Cornerstone Research provides economic and financial consulting and expert testimony in high-profile litigation, investigations, and regulatory matters. We support clients with rigorous, objective analysis that is grounded in real-world data, state-of-the-art research, and case precedent. We are passionate about exceeding our clients' expectations and we saw early on that data science would be essential to upholding our standard of excellence.

That is why we created the Data Science Center, our hub of data science expertise. The Data Science Center is a

pioneer in our industry in applying modern techniques such as artificial intelligence, machine learning, and text analytics to supplement more traditional econometric analyses. Our mission is to maintain our leadership position by continuing to set the standards for technology, data science, and data engineering.

What value does the Data Science Center bring to the firm's casework and for clients?

The Data Science team brings value in four main areas: increasing efficiencies, unlocking the potential of data, providing scalable/bespoke solutions, and supporting defensible results. Let me address each in turn.

One example of the efficiencies we provide to clients is our use of IBM Netezza, which provides speeds 20 to 2,500 times as fast as conventional analytical platforms. The increased computational speed of Netezza often directly translates into key strategic benefits and outcomes. These may include:

- Timely identification of data quality deficiencies in produced data
- The ability to conduct the numerous necessary iterations, sensitivities, and robustness checks for an expert analysis
- Quick turnaround on urgent requests, even when dealing with large datasets that typically require substantive computational processing times

Regarding unlocking potential, I am referring to our ability to digest and analyze large-scale data that opens up new lines of investigation. We are able to conduct analyses and pursue ideas that would have been infeasible a few years ago—or would have taken a massive amount of resources to produce.

For example, in the *T-Mobile/Sprint merger*, our experts analyzed how consumers choose wireless carriers and



how wireless carriers compete. This analysis used highly granular data comprising billions of data points on when, where, and how consumers used their mobile phones. Data Science was able to conceptualize and create an algorithm to efficiently categorize this information in space.

With respect to scalable/ bespoke solutions, the Data Science Center focuses on delivering right-sized solutions. For example, we can develop bespoke analytic approaches when appropriate or use scalable tools to help our teams automate analysis. So, in addition to providing clients with efficient and secure large-scale data analytics, we are also able to offer customized applications informed by years of experience.

By “defensible results,” I am referring to our experts’ ability to demonstrate and communicate analyses and findings using both traditional methods and more cutting-edge technology supported by our Data Science Center.

Big data is a term that we have been hearing for several years. How is Cornerstone Research’s Data Science Center equipped to handle the challenge of an increasingly data-driven world?

It is an interesting challenge that we have been able to meet successfully. What makes this work especially challenging is not only that the magnitude of data we are asked to process and analyze has increased exponentially, but there has also been an explosion in the kinds of formats we are handling.

As to how we deal with substantial volumes of real-time and historical data, we have heavily invested in secure, on-premises analytics infrastructure with massively parallel processing capabilities, such as IBM Netezza. We regularly work on cases with hundreds of billions of data records. We are also experienced in leveraging cloud computing capabilities for surge storage or compute capacity. Our team

of programming specialists and data engineers ensures that we can conduct large-scale data analytics efficiently and effectively in a fraction of the time it used to take.

The second biggest challenge with big data is that we often will be asked to work with client data and other private and public sources from a wide variety of platforms and incompatible formats, which must be processed to provide reliable data. Our Data Science capabilities allow us to help counsel manage the discovery and data production process efficiently and effectively. We also work with clients to extract information in anticipation of the analytical needs of subsequent phases of work as well as in response to direct requests from regulators or litigants.

For example, in a high-frequency trading matter, the Data Science Center worked with several stock exchanges to access more than 200 TB of data in order to determine the best protocol to access these large and complex datasets, as well as to identify the relevant data subsets that would be used for the necessary analyses.

Big data also increasingly encompasses more than traditional structured data that comes in rows and columns. Our experience with AI and machine learning is valuable when analyzing unstructured data, including documents and text, images, video, and audio.

How can AI (artificial intelligence) and ML (machine learning) support expert testimony?

AI-based systems substitute human decisions with data-driven ones. This can reduce subjectivity and error when processing large volumes of complex information. We utilize AI and ML to drive automation of increasingly complex tasks and unlock new approaches for analysis, including using both supervised and unsupervised learning.

Our machine learning capabilities are enhanced by our in-house graphical processing units (GPUs).

We have invested in a number of off-the-shelf and proprietary software tools, packages, and data pipelines to facilitate efficient analysis.

GPUs provide computational speeds that exceed those of even the fastest central processing units (CPUs). For example, in antitrust matters, we often need to calculate the distance between all suppliers and all consumers (coordinate pairs). Migrating this computation from CPUs to GPUs enables us to calculate distances between nearly 100 million coordinate pairs per second.

Social media and big data are the most prominent trends of the 21st century. How is the Data Science Center helping companies keep pace with these intertwined technologies?

With their vast user base that eclipses traditional media, social media platforms offer rich sources of data that multiply at dizzying speed. In litigation contexts, knowing how to effectively navigate, collect, and characterize such huge amounts of data is crucial. In addition to our deep familiarity with social media data sources, our experience with AI and ML tools equips us to assess the relevancy and relative prominence of content and contributors. This is a fast-growing area of data and the insights these sources provide can be crucial in supporting expert analyses of text, content, and sentiment.

What about some examples to illustrate that?

For large-scale analysis of Reddit subforums, commonly known as subreddits, we built web data pipelines and automated approaches, leveraging ML to score a post's textual/context-driven relevance to topics of interest and characterize the prominence of a given post relative to other posts in the subforum.

In connection with *In re Facebook Inc. IPO Securities and Derivative Litigation*, we employed advanced language models to effectively distinguish homographs in tweets and generate features for an ML classifier. This framework facilitated the reliable and scalable detection of public awareness of alleged material omissions prior to required disclosure.

Lastly, we have extensive experience with online consumer reviews of products and services—among the most intriguing (and demanding) social media data. These reviews can be the subject of litigation, but if employed appropriately, they can also provide a valuable source of real-world data. We're skilled in evaluating these distinctive data, including assessing the relative importance of product features, changes in customer sentiment over time, and fraudulent reviews.

Can you talk a bit about Cornerstone Research's investment in the Data Science Center? What technology and training supports its work?

In litigation, we often deal with sensitive client information, so we invested heavily in secure infrastructure, including high-performance and high-throughput analytical servers and storage clusters. Our analytical infrastructure is on-premises, meaning client data are not exposed to the web.

We have also invested in a number of off-the-shelf and proprietary software tools, packages, and data pipelines to facilitate efficient analysis. For example, when working with documents, we utilize tools to add high-quality text layers to documents, quickly extract tabular data, and develop tailored approaches to extracting other key information.

Finally, we have invested in people. We have exceptional data scientists and practitioners with many years of experience across a large number of different clients and projects. Mike DeCesaris, who is the vice president of the team, has a background in economic consulting and

computer science, which puts him in an ideal position to navigate the profound transformations that litigation and expert testimony continue to undergo in regard to data.

How do Cornerstone Research's in-house experts and network of outside experts, which include leaders from academia and industry, work with the Data Science Center?

Cornerstone Research's testifying experts are at the forefront of litigation trends, industry innovations, and academic research. In turn, our experience with implementing sophisticated data science techniques supports these experts in their analyses. Experts appreciate the fact that we bring such a deep understanding of AI and machine learning to automate complex tasks and develop analytic approaches to supplement traditional econometric and statistical methods. For example, Data Science staff applied machine learning approaches to healthcare risk adjustment models, which explained approximately twice as much variation in claims data as the status quo linear regression model. ■

The views expressed in this article are solely those of the speaker, who is responsible for the content, and do not necessarily represent the views of Cornerstone Research.



Greg Negus has more than thirty years of experience in professional services firm management. As the chief operating officer at Cornerstone Research, Mr. Negus is responsible for leading the firm's corporate and administrative functions, including finance, information technology, human resources, talent, and marketing. He also manages the firm's facilities and office administration operations. Mr. Negus serves on Cornerstone Research's executive committee and its pro bono committee, among others. Reach him at gnegus@cornerstone.com