Booz | Allen | Hamilton

Harnessing Big Data and Analytics for Precision Medicine

A New Approach for a New Age

by Dr. Kevin Vigilante

Prepared exclusively for Organization Name

TABLE OF CONTENTS

A NEW APPROACH FOR A NEW AGE IS NEEDED TO UNDERSTAND AND TREAT DISEASE

By Dr. Kevin Vigilante

Vast amounts of data are becoming available for precision medicine, giving us the potential to make quantum leaps in our understanding of disease and how to treat it. But so far, we've had only a limited ability to use all this big data—much of it is scattered in countless isolated databases that have been difficult to bring together to get the big picture.

That's changing. A new approach, developed by Booz Allen, is now making it possible to integrate a virtually unlimited amount of precision medicine data—and put it directly into the hands of researchers and other users.

The advent of precision medicine marks the first time that the life sciences have truly entered the realm of big data. While other sciences, like meteorology and astronomy, have used big data for some time, it's now precision medicine's turn. We're gaining access to an ever-growing tsunami of information, including genomic and other biospecimen-derived data, survey and demographic data, clinical and insurance-claims data, mobile and implant data—the list goes on.

Alan Kay, the pioneering computer scientist, has said that big data is really about "big meaning." And if we are to realize the promise of precision medicine, we must be able to find that big meaning. Yet with traditional computing methods, there are major obstacles at every turn.

For example, a large portion of the data essential to precision medicine is locked up in conventional relational databases, in a profusion of formats. These databases typically were built to achieve specific purposes, or answer narrow types of questions, and cannot easily be integrated. As a result, only a relatively small amount of data can be brought together to solve a precision medicine problem.

A second problem is that with relational databases, we have to know in advance the kinds of answers we expect to find. Researchers must first pose hypotheses about what the data might say, and then custom-build the databases to test the hypotheses. Yet an essential part of precision medicine is going beyond what we know—or think we know—by finding the kinds of hidden correlations and patterns that reveal the "unknown unknowns."

We need to let the data talk to us, so to speak. Conventional databases don't do this well.

In addition, much of the data we might use for precision medicine is "unstructured," and cannot easily be formatted for conventional relational databases. Examples include images and text, such as doctors' and nurses' notes, and social media posts. Without these data sources, precision medicine can progress only to a certain point, and no further.

Conventional methods are also of limited help in overcoming one of the greatest obstacles to precision medicine, and that's data sharing. The puzzles of precision medicine cannot be solved by a single government agency, life sciences company, or university. It requires what might be called a "megacommunity," with all of government, business, and society pooling their data—and often their expertise—in a collaborative effort.

And yet traditional approaches to data cannot provide the necessary security and privacy protections that each stakeholder needs. Once again, precision medicine hits a brick wall.

Still another challenge is that with conventional methods, researchers without specialized training in IT don't have direct access to the data. They must go through data scientists and other IT specialists in a laborious process to ask questions and get back answers. But with the explosion of data and analytics in virtually every corner of society, these IT experts are in high demand and short supply. There are simply not enough of them available to help researchers explore many of the most promising avenues of precision medicine.

The ultimate problem is that current computing approaches weren't built for big data—they were developed when data was "smaller" and more manageable. The traditional methods worked well for many years, but they are not up to the task of addressing big-data challenges like precision medicine.

APPLYING LESSONS FROM THE INTELLIGENCE COMMUNITY TO PRECISION MEDICINE

We recognize that to realize the promise of precision medicine, we need an entirely new way of thinking about big data. We must be able to quickly and easily bring together massive amounts of information—unstructured as well as structured—and make it readily available for analysis.

We must be able to let the data talk to us, so we can find those valuable correlations and patterns. We need new ways of ensuring security and privacy, so that the diverse stakeholders of the precision medicine megacommunity will freely share data. And we must be able to put the data and analytics directly into the hands of researchers.

These goals are within reach. A new approach to data and analytics—initially developed through an ongoing collaboration between Booz Allen and the U.S. government to combat terrorism and other threats—is now being applied to precision medicine.

It arose from a thorny problem facing intelligence analysts. They were collecting massive amounts of intelligence data from a broad range of sources, and yet had only a limited ability to bring it together. We helped the government develop a new approach, one expressly designed for big data. It has enabled intelligence analysts to use all of their data to see the bigger picture. It has also been used by U.S. military in Afghanistan and Iraq for various intelligence purposes, such as identifying militant groups responsible for roadside bombs.

This new approach has been moving into the larger government and business communities, most notably the health and financial sectors. For example, data scientists at Booz Allen, working with a hospital chain in the Midwest, brought together and analyzed large amounts of data on patients with sepsis, the life-threatening response by the body to infection.

They found previously hidden patterns in thousands of patients' vital signs, indicating when a patient was about to go into severe sepsis, a more dangerous condition. As a result of the findings, hospitals in the chain immediately started monitoring sepsis patients for such red flags, so that doctors could implement their protocols earlier and save lives.

We believe our new approach is ideally suited to solving the kinds of big-data problems that threaten to hold back progress in precision medicine. It is about much more than technology. It really represents a new mindset—a reimagining of how information can be used to further medical discovery.

THE PARADIGM SHIFT OF THE DATA LAKE

At the heart of our approach is what we call the "data lake," a new way of storing and managing data. Data is no longer locked in limited, isolated databases. Instead, all the available data is consolidated into a single pool, or "lake." It is both stored and analyzed in the cloud, using networks of computers.

What makes this new approach possible is the way the computer finds the data. With a relational database, each piece of data is assigned a location based on rows and columns, as with a spreadsheet. Because the data has to be painstakingly formatted, this method only works well with relatively small amounts of data.

The data lake solves this problem by identifying the data in a way that doesn't rely on rows and columns. Instead, as each piece of data is put into the data lake, it is "tagged" with accompanying details that can be used to locate it. For example, a piece of patient information, such as genomic data, can be tagged with other information about the patient, such as age, medical condition, medications, income, etc.

Tagging works for big data because it is much faster than conventional formatting methods. Data that would have taken days to format for relational databases can be tagged for the data lake in a matter of minutes.

This new method has several important advantages. First, we can now integrate a virtually unlimited amount of information, from any number of data sources. All of the available data can be analyzed for insight—all at once. This opens the way for entirely new realms of inquiry in precision medicine.

In addition, the data lake easily accepts unstructured data, which can now be searched for insight side-by-side with other types of data. Our approach uses advanced analytic methods to prepare and tag the data for the data lake. For example, we use machine learning and computer vision to identify the content of images. Because this can be done in an automated process, it eliminates the labor-intensive manual work required in conventional methods.

We also use natural-language processing and other techniques to make clinicians' notes, research papers, social media content and other text available for the data lake. These methods have been used by the intelligence community to understand the ideas contained in covert messages. We are now applying them to precision medicine.

LETTING THE DATA SPEAK FOR ITSELF

By removing the need for relational databases, we no longer have to decide in advance what kinds of questions we want to ask. Using the tags, we can pursue any line of inquiry, looking at all the data in the data lake.

If we decide to ask a different type of question, we don't need to tear down the database and laboriously custom-build another. We simply go back to the data lake and reframe our inquiry.

Our advanced analytics allow us to take this a step further. Using the tags in the data lake, the analytics let the data speak for itself, revealing the kinds of hidden patterns and connections that can be so valuable in precision medicine. Once researchers find potential insights, they can test them with a traditional hypothesis approach. But the key here is that they don't have to start with the hypotheses. They don't have to suppose what the answers are before they even ask the question.

One might think that consolidating large amounts of data in the cloud makes it less secure. But with the data lake, the opposite is true. With conventional databases, it is often difficult to determine who is accessing the data, and what they do with it across their systems. The more that organizations share their data, the more their risk grows.

The data lake solves this problem. Each piece of information in the data lake is put into an individual cell, along with its tags. In addition, each cell is embedded with visibility controls that tightly govern who has access to that specific piece of data, and under what circumstances.

This is essentially a security filter that limits—cell by cell—what someone can see in the data lake. And it ensures that all compliance regulations, standards and legal restrictions are strictly applied to each piece of data.

This granular level of security and privacy is critical to bringing together the precision-medicine megacommunity, by giving each stakeholder confidence that it can share its information.

BOOZ ALLEN: YOUR ESSENTIAL PARTNER FOR PRECISION MEDICINE

Booz Allen's approach also features innovative visualization tools that give people who are not computer experts direct access to the data and analytics, without the need for intermediaries. Precision medicine researchers can ask questions of the data and look for patterns and connections—using plain English, as easily as doing a Google search.

While data scientists may still need to help with particularly complex inquires, researchers can do much of the work on their own. They can freely explore the data, pursuing new ideas and following hunches wherever they might lead.

Democratizing the data and analytics goes beyond technology. As part of our approach we help organizations build a spirited culture of analytics by focusing on elements such as collaboration, experimentation, and buy-in. And we provide training that shows non-experts how to use data and analytics to get real-world precision medicine insights. One of the most important lessons we've learned as management and technology consultants is that problems aren't solved by tools alone. It's how the tools are paired with people that counts.

We've been using this kind of dynamic approach to solve complex problems for more than 100 years. We have applied it successfully at every federal agency and department, and across analogous commercial sectors.

Our business is built on our understanding of how to help clients think through their goals, and then achieve them—often in groundbreaking ways.

Our expertise in data science and advanced analytics is unparalleled. Our data scientists—many of them among the earliest members of the profession—continue to invent breakthrough technologies and analytic approaches, transforming how business and government operate.

We now have nearly 600 data scientists—one of the world's largest such teams—whose expertise reaches across every federal sector, including health, finance, transportation, environment, defense, intelligence, and homeland security.

We're also committed to helping healthcare and life sciences organizations across the private and public sectors navigate their rapidly changing environments and complex markets to drive more effective treatment and approaches. Overwhelming amounts of data present both opportunities and challenges for clients, offering new and innovative ways to discover insights while creating an urgent need for privacy and security.

Through the combined power of our management consulting, data science prowess, and life sciences experience and expertise, we are helping organizations harness big data and advanced analytics to reach new levels in precision medicine.

About Booz Allen

Booz Allen Hamilton has been at the forefront of strategy and technology for more than 100 years. Today, the firm provides management and technology consulting and engineering services to leading *Fortune* 500 corporations, governments, and not-for-profits across the globe. Booz Allen partners with public and private sector clients to solve their most difficult challenges through a combination of consulting, analytics, mission operations, technology, systems delivery, cybersecurity, engineering, and innovation expertise.

With international headquarters in McLean, Virginia, the firm employs more than 22,600 people globally and had revenue of $5.41 billion for the 12 months ended March 31, 2016. To learn more, visit BoozAllen.com. (NYSE: BAH)