THE NEXT GENERATION OF DIVERSITY METRICS: Predictive and Game-Changing Analytics

Edited by Brian Baker, General Manager, Diversity Best Practices
and Michael Collins, Leader, Member Experience, Diversity Best Practices

Predictive and Game-Changing Analytics
What if we could predict the future of diversity and inclusion by looking at the numbers?

Most diversity practitioners would likely have a wish list of things they would feel we could accomplish if we could predict behaviors of the people in our organizations. And given today’s robust electronic world, a vast data trail capturing people’s behavior is inevitable. A commonly used term to describe that trail is “Big Data.”

Some of the more stunning data about Big Data: Every day, people around the world create 2.5 quintillion bytes of data. (Quintillion: That’s a 1 followed by 18 zeroes.) This data is gathered from a wide variety of sources, including climate-based sensors, posts to social media sites, digital pictures and videos, purchase transaction records, cell phone GPS signals, and much more.

The explosion of data is so recent that 90 percent of the data in the world today was created in only the past two years. Enterprises alone gather terabytes (one million million), and even petabytes (1,000 terabytes), of content. The data moves so rapidly that sometimes even two minutes is too late, particularly when it comes to time-sensitive processes such as catching fraud. Data comes in multiple shapes and sizes—it can be structured and unstructured—ranging from clicks on a website to uploaded videos. Putting various types of data together can yield new, powerful insights.¹

It’s one thing to sit back and analyze an overwhelming mass of data about what happened in the past. It’s quite another to make this data speak to us about the future. Angelina Jolie, for example, was able to make a brave decision to proceed with a double mastectomy based on data that told her that breast cancer was in her future—data that was not available only a handful of years ago.²

At first blush, Big Data may not seem to be the bailiwick of diversity practitioners. That’s for the IT or marketing people, we may tell ourselves. But increasingly, capturing data and being able to interpret what people and organizations are going to do next (predictive analytics) is proving itself germane to diversity and inclusion.

Predictive analytics will soon offer the make-or-break evidence needed to support every business case, every new project proposal. For diversity practitioners, predictive
analytics offers more: A powerful tool to be smarter about inclusion efforts—which ones to ditch, which ones to double down on, which ones to invent.

The upside to pursuing predictive analytics is high. Data will, increasingly, be gathered, assessed, and used to predict. The time is coming—rapidly—when leadership acumen will be dulled if we can’t find a way to use predictive analytics to guide and shape our strategies and programs.

**Mouse Tracks**

Let’s rewind and examine how predictive analytics is showing up in every day life—and then how we can apply these new realities and tools to the work of advancing diversity and inclusion.

Our journey starts with observing the digital tracks of a guy we’ll call Roberto as he surfs on his computer within the space of one hour. He’s eating lunch in his home office, making an Amazon.com purchase after having just completed a Google search about the headphone set he is buying. Moments earlier, he updated his Netflix movie queue and popped last week’s watched movies into his mailbox for return. Once he completes his Amazon purchase, Roberto plans to review e-mails, check his Twitter feed for interesting items he can reweet, and then resume working. But before he can get to his e-mail, he takes a call on his cell phone.

You may already have some sense of the data trail Roberto has left in the course of this hour. If you shop at Amazon, you frequently see “things you might like” and similar recommendations that are based on items you’ve purchased and the products you’ve studied on the website, but not purchased. Likewise, if you’re a Netflix customer, right-on-the-money movie suggestions, just for you, pop up every time you visit the site. You may know those suggestions are a great match because they are based on movies you’ve watched, and rated highly, in the past.

But these sprinklings are only the beginning of Roberto’s data trail. His Google search activity has been recorded, and added to Google’s vast body of user statistics that can be modeled to predict future trends and user needs. E-mails on which Roberto has clicked are tracking interest, as are any websites Roberto is drawn into in the course of perusing e-mail. His every click, let alone tweet on Twitter, is recorded. And Roberto’s location—should we ever need to verify it—is being tracked by the nearest cell phone tower. Within one hour, Roberto has left countless clues to his identity, his preferences, his interests, and his behaviors. And Roberto will continue to add to that data trail throughout his day.

What happens with this data as it is accumulated? While the National Security Agency is looking for patterns that it may use to alert Homeland Security, businesses are using the data for a variety of marketing-related activities, including better segmenting and targeting of its audiences and improved service provision. As we’ve seen, Netflix, Amazon, and also iTunes use the data to accurately recommend songs or movies you are likely to enjoy. They aren’t doing it just to be nice—although it certainly adds to the perception that they understand you and are uniquely equipped to meet your needs. More to the point, these providers are enticing you to buy more by offering you things they know you’ll like based on your past behaviors.

Marketing is only the tip of the iceberg when it comes to uses for this data. As we know, big data is everywhere. One of the newest, more innovative uses is to monitor health trends. Given the speed and volume of Twitter, Facebook, and other social media outlets, health organizations including the U.S. Centers for Disease Control and Prevention (CDC) are using social data to monitor trends around flus and other epidemics.

There’s Google Flu Trends, which glean flu-related insights from social media and correlates the data on regional maps. MappyHealth is a similar provider that can spot where a flu outbreak is manifesting and can predict where the next outbreaks will take place.

Let’s examine how it works. As of the two weeks ending July 31, 2013, MappyHealth had gathered more than 2.6 million tweets that met its parameters regarding illness-related conditions. Those tweets offer a snapshot regarding which illnesses are most prevalent among social media users, which can then be correlated to the general population. Only about 1 percent of those tweets could be tied to a location—but that still offers more than 52,000 postings contributing to the geographic snapshot of an illness. Factor in the fact that 1.7 million tweets can be connected to individuals who’ve associated a location with their profile, and geographic data about illnesses becomes that much more comprehensive. Google can add search data to this mix—whenever a user conducts a search using the right key words (such as “stomach flu”) the data is captured and tracked, across more than 25 countries.

MappyHealth can also connect illness-related tweets with events associated with the tracked terms, which provides information that can be used for proactive steps or to evaluate recent efforts. For example, in 2013, between May 2 and May 8, the website reported three spikes in the number of tweets concerning the H5N1 virus, commonly known as the bird flu. Between May 2 and May 4, the number of tweets on this topic spiked twice, with the first spike occurring on May 2 showing an increase of 1,950 percent over just days earlier. The second spike on May 4 was not as dramatic, but still significant. Further analysis of this information revealed that an event had occurred around the same time as these spikes. Researchers at China’s National Avian Influenza Reference Laboratory at the Harbin Veterinary Research Institute released a study that indicated the bird flu could evolve and be passed on to mammals. The third spike in bird flu tweets on the MappyHealth site was related to the positive diagnosis of the bird flu in the Crested Myna, a starling species native to southeastern China and Indonesia. Tweets for the third spike were primarily made in Malaysia.

Two things come to mind in looking at MappyHealth’s experience with the bird flu. First, the research reports were dramatic and the channels to report on the findings were successful—people tweeted about it. Second, how might health or wildlife officials around the world use this or similar information in preparing for, say, an influenza outbreak in people or take proactive actions to protect domestic birds before the virus could affect the food chain. Because MappyHealth includes influenza in its tracking algorithm, businesses could proactively offer influenza vaccinations for its employees prior to a local flu outbreak, and thereby protect employees from illness and the business from productivity downturns.

**Predictive Analytics for the Diversity Practitioner**

Several major companies and progressive organizations are currently looking at how predictive analytics can influence some functions in human resources. But don’t wait for the Big Data trend to come to you. Diversity practitioners who jump now will be leaps and bounds ahead of their peers who hang back. For diversity practitioners, it’s time to get to know the IT folks better, not to mention your employer’s various available data sources. Does the business conduct an engagement survey? Are online behaviors, such as intranet clicks, trackable? What might you be able to do with those pieces of data?

Diversity leaders can approach predictive analytics from either end. They may begin by understanding available data and tools, and identifying behaviors worth assessing, or a particular challenge or initiative or program might merit making an investigation to see if predictive analytics can offer a possible solution. The best place to start is with human resources, where talent analytics has been on the table for some time already, and as a result presents a huge opportunity for predictive analytics around diversity and inclusion. It’s also one of the areas where companies may already have a level of comfort for this new tool.

For many employers, human resources or their people management systems—recruitment, benefits, compensation, training and development, and others—represent one of their largest expenses. The workforce represents an essential component of product or service delivery. Tightly controlled management of workforce costs represents the Holy Grail for most enterprises. Predictive analytics can help any business ensure a vital,
DIGGING INTO DATA: How Predictive Analytics Works

Applying formulas and equations to raw data can help predict what will happen in the future. As described on the Predictive Analytics World website, “Predictive analytics is business intelligence technology that produces a predictive score for each customer or other organizational element. Assigning these predictive scores is the job of a predictive model which has, in turn, been trained over your data, learning from the experience of your organization.”

The core component of predictive analytics is the predictor. This represents a single value that is measured for each person—a single click into a company’s intranet, a single e-mail opened, the time period used before a person made his or her health plan choices, for example. Each one of these predictors can be used to anticipate future behavior.

For example, if a person waited for three weeks, until just before the deadline, to make healthcare plan choices, it is more likely than not that he or she will require a similar amount of time to make choices next year.

To create a predictor one simply needs to apply a defined behavior (making health plan choices, for example) with a data set (amount of time taken for employees to make their health plan choices).

Using one predictor provides helpful input. Combining predictors results in more powerful data. When you combine predictors, you’ve made a model. Models can be simple, or they can be complex. Consider the combinations. For example, combining the number of times (if at all) a person accessed health plan materials online with the amount of time a person required to make health plan choices makes for a better predictor. Here’s where it gets even more interesting, particularly for diversity practitioners. Things like race, sexual preference, gender, and ability can be predictors. It’s one thing to predict that a person will take three weeks to make health plan choices because he or she required that amount of time last year. It’s another thing to say, as an example, that Asian women in the aggregate who accessed online health plan materials made their health choices in only two days.

The typical predictive model used in consumer marketing is highly complex, and often combines dozens of predictors. The more complex a model, the better, richer, and more accurate the predictions will be.

Finding the best predictive model is the key to generating accurate predictions. In the case of predictive analytics, data mining technology is unleashed on your data—for organizations, this might be their comprehensive workforce data—to automatically build a predictive model based on that company’s unique data. The technology literally learns from past behaviors. Using our health plan example, it reviews whatever records the organization has captured over many years to uncover patterns and incorporate its findings into the predictive model. As stated by author and analytics expert Eric Siegel, “Predictive modeling software has computer science at its core, undertaking a mixture of number crunching, trial, and error.”

The following graphic illustrates the process.

It all seems wonderful, relatively simple, a bit high-tech, and fully automatic. Yet, expertise and experience are required to ensure a model that is sufficiently—but not excessively—complex.

The person requiring the data must be involved in shaping and reviewing the model, to ensure the best, most accurate predictions. Part of shaping the data involves determining just what is to be measured. Generally, business leaders need to team with analytics people to identify an appropriate goal to measure. The challenge is that your ideas regarding what to measure may be plentiful, but not all can be captured using data analytics. Further, even when a large set of behaviors can be predicted with good analytic data, not all of those behaviors are business actionable or worth assessing from a business standpoint. Perhaps, for example, we can pinpoint which evening hour is the most popular, among a key population, to peruse healthcare plan data from home. Great, but just how useful will that particular piece of data be?

By contrast, it might be well worthwhile to identify whether certain populations are more likely to want to peruse health care plan data at home. Do enough workers prefer reading through their plans at home that it’s worth providing plan detail in an easily portable format? As the goal becomes clearer, details can be added. What specific populations are more likely to want to review the plan information at home? Which would prefer only reviewing the information at work? Are other family members involved in reviewing the information?

Once you’ve chosen a clear, well-defined, detailed goal that you know is actionable, it’s time to test the model. You may discover that the results are not strong enough, suggesting a need to better or differently define your goal. Or, your trial run may reveal underlying issues—the data captured is helpful, but the timeframe for capturing the data is too late in the process. Capturing predictive data is never a finished job—the equation likely will change, via either tweaks or larger scope shifts, on an ongoing basis. The process is fluid. You’ll better hone in on behaviors to measure once the data rolls in.
“Once these findings were incorporated into the recruiting process, the company increased revenues by $4 million in six months.”

Predicting Which Candidates Will Be Most Likely to Succeed
Consider the case of a financial services company that, for years, believed employees who’d attended highly ranked universities and scored good grades made the best performers. The assumption seems logical; it wouldn’t be surprising if many organizations hold similar beliefs. But in reality, this company’s data suggests otherwise. Where a person went to school and what grades he or she obtained were less likely to predict success in the workforce than other, perhaps surprising, factors, including a lack of typographic errors on a resume, experience in selling (in this case, selling real estate or automobiles), an ability to respond to vague instructions, and experience in planning daily activities and shuffling priorities. Once these findings were incorporated into the recruiting process, the company increased revenues by $4 million in six months.

Predicting Which Types of Managers Were Most Likely to Retain Talent
Google, committed to incorporating Big Data into its HR practices, found similar results from its own data. College GPAs and SAT scores were, at one time, a key element in screening candidates, but when Google found that these two numbers did not predict workforce success the data was eliminated from candidate screening processes. But Google has taken its analytics a step further. It recently did a study to determine if, and how, good managers make a difference. Prasad Setty, head of Google’s people analytics group, says, “Through various methods, we found positive relationships between good manage-

ment and retention and the performance of teams. We then conducted double-blind interviews to identify the key behaviors exhibited by our best managers. We found eight behaviors that make a good manager and five pitfalls to avoid. These are now integrated into our manager-training programs and coaching sessions, and teams provide feedback to managers on these behaviors to help them understand where they’re doing well and where they can get better. The vast majority of our lower-rated managers have improved as a result.”

Improving Talent Management
Another employer, PNC, sought to incorporate line-executive input in identifying ways to improve its talent management approach. After interviewing those line executives, the company’s analytics team developed a list of questions and hypotheses to test. That output will be used to build various data sets that PNC can use to develop talent. According to Jay Wilkinson, PNC’s HR vice president of analytics, “This is where HR has the chance to prove itself. Better than coming to [business leaders] with tired best practices, we’re asking them how they define success specific to their business, and that provides the context for our analysis and recommendations.”

Predicting Who Is Most Likely to Quit
Hewlett-Packard is another employer making strides in tapping talent analytics. The company’s massive analytics department houses 1,700 employees in Bangalore, India alone, looking at sales, marketing, finance, supply chain, and HR. Gitali Halder, who heads HP’s Bangalore HR-focused analytics team, partnered with fellow HP consultant Anindya Dey to compile large quantities of performance and related data about HP workers to serve as learning material. They looked at salaries, raises, performance ratings, and job rotations, adding, for each individual, whether the person had quit. The result? A data pool that, if tapped effectively, can predict the various factors that make someone most likely to quit their job. In HP’s case, some of the factors associated with people quitting are just what you’d expect. Higher salaries, higher performance ratings, and more promotions generally meant a lesser tendency to leave. Interestingly, HP also found that more job rotations meant a person was more likely to hang around. Further, although promotions generally decreased a person’s tendency to resign, in certain departments it actually increased that tendency. It’s these finer points of data that hold high value for HP, or for any organization. Being able to tweak specific issues within specific departments to increase retention is an amazingly valuable capability.

Diversity leaders must follow in HR’s footsteps when it comes to talent analytics. If your company already is capturing and using talent analytics, you need to identify ways to use that data in the service of diversity and inclusion. Google used analytics to completely upend its HR function and lead the pack in its use. From the name of the function—people operations instead of human resources—to how it uses data and analytics to inform every area of responsibility, Google has replaced traditional HR practices with ones that are based on people analytics or data-based decision making.

Let’s look at three areas where Google’s people analytics intersects with diversity issues. For recruitment, the company uses a hiring algorithm that predicts which candidates have a high probability of success once onboard. One of the results of using this protocol is that Google has shortened its recruitment processes, making the company more attractive to potential new hires. The recruitment analytics also analyzes rejected resumes to identify any top candidates who might have been overlooked during the initial resume review. The miss rate was only 1.5 percent, but Google revisited some of those candidates and ended up hiring them.

Next, let’s look at Google’s use of people analytics in retention. Because of the nature of their business, Google takes retention so seriously that it developed a special algorithm to evaluate and predict which employees are most likely to leave. With this kind of information in hand, managers are able to step in before an employee leaves and offer personalized incentives for the employee to stay.

Google also uses predictive modeling to anticipate upcoming people management challenges and opportunities as well as for workforce planning management. By consistently asking “what if,” succession plans, leadership development and performance management can be influenced by its predictive model.
As a diversity practitioner, consider the potential benefits of linking the hiring algorithm with the recruitment of candidates representing various dimensions of diversity, along with the chance to revisit overlooked high potential resumes. Or perhaps it would be helpful to analyze retention data to make sure the women and employees from various underrepresented demographic groups are approached with personalized incentives to stay before they are allowed to slip through the cracks and out the door. And think about how predictive analytics could level the playing field in the succession-planning and promotion or leadership-development process.

Not long ago, Google's people analytics team used their models and algorithms to look at the diversity of its workforce. They sought to identify the root causes of weak diversity recruiting, retention, and promotions (especially among women engineers). Using its data-based decision making, efforts to address its diversity and inclusion challenges achieved dramatic—and measurable—results in hiring, retention, and promotion. Google has harnessed the power of predictive analytics to transform its HR processes from start to finish. Diversity practitioners are well advised to consider how to connect diversity issues with the power of analytics to do the same.

**Data Mining + Predictive Analytics = D&I's Future**

The ground is shifting underneath our organizations. The future of diversity and inclusion promises to be infused with data mining and predictive analytics models. These tools represent another evolution of business operations and are becoming a competitive necessity for any organization—from marketing, IT, risk management, and customer service, to human resources and diversity management. Below are several examples of how three major companies tap into the power of Big Data to recharge, reenergize, and renew the D&I function. While some companies are further along this journey than others, all of these organizations recognize that Big Data is here to stay, and diversity executives are working to leverage it on behalf of their diversity and inclusion strategies.

**Whirlpool: Using Predictive Analytics to Give Added Legitimacy to Diversity and Inclusion Goals**

Whirlpool is among the businesses beginning to use predictive analytics in the service of diversity and inclusion. Predictive analytics results are broadly used throughout the organization, because initial efforts at data modeling have sparked deeper conversations. Preliminary insights and implications from predictive analytics are leading to an interest in doing more. For Whirlpool, predictive analytics help provide a baseline for risk assessment, learning and development efforts, talent intake, and retention strategies around diversity and inclusion. According to Ingrid Jacobs, diversity & inclusion senior manager, current modeling analyses have helped steer Whirlpool away from a “myopic focus on intake, while ignoring potential effects of retention risks and advancement challenges for diverse populations.”

Those same predictive models have helped the diversity team explain to HR and the business that continuing on their current trajectory would not yield results helpful to Whirlpool. Thanks to predictive analytics, conversations are shifting from reactive debates to proactive consideration.

Whirlpool specifically tapped big data to support its strategic objectives around increasing leadership diversity, both among actual leaders and within the leadership pipeline. To gather the right data, Whirlpool took a deep dive into diverse representation in attraction, retention, and advancement.

According to Jacobs, “We pulled historical data on hiring, promotions, attrition, employee engagement scores, performance ratings, retention risk, recognition, and learning and development. Using this data we created a diversity scorecard showing current state and—thanks to predictive models—potential future state. The scorecard and models were presented to senior leaders to develop action plans.”

Whirlpool, like many organizations, has faced some challenges in using predictive analytics for diversity and inclusion efforts. Most notably, the “lack of data, lack of resources, lack of best practices and case studies, the limits of analytics within diversity and inclusion mirror the limits for the greater HR function, said Jacobs.

Despite these challenges, Jacobs continues to believe in the value of data-based decision making. And with predictive analytics supporting the business case, Whirlpool now is seeing a culture of accountability for results around its various diversity and inclusion efforts. Further, the company has gained a better understanding of how current practices can influence future business outcomes. Even better, using predictive analytics has “provided legitimacy to the goals of diversity and inclusion by quantifying the effects of status quo behaviors,” said Jacobs.

In the future, Jacobs will continue to pursue more efforts that tie business results directly to diversity and inclusion efforts.18

**Walmart: The Predictive Analytics Veteran**

Thanks to its scope and size, Walmart has been “into Big Data before Big Data was cool,” said Donald Fan, senior director, global office of diversity. The company processes 240 million transactions every week, supports and maintains a half-million network devices, operates 38,000 servers, and is building the largest private cloud of any business to date. Big Data and predictive analytics help Walmart forecast its customer’s needs, enhance their experiences and better serve them through multiple channels, whether they are online, using a mobile device, or at a store.

Consequently, Walmart applies predictive analytics to its workforce via internal and external channels. Internal data helps to spot talent gaps, monitor growth trends, identify future talent needs, and design targeted initiatives to keep the company moving toward its next level of excellence. Meanwhile, external predictive analytics help Walmart monitor its progress compared to competitors or admired businesses, keep tabs on best practices, and stay ahead of potential challenges down the road.

The company’s Diversity Goals Program holds the management team accountable for elevating diversity and inclusion standards throughout the company. The program impacts pay, with up to 15 percent of officer and manager bonuses at stake, based on diversity goal achievements. More than 67,000 managers across Walmart’s U.S. operations are accountable for meeting diversity goals. Consistent data mining, reporting, and analytics help monitor progress toward goals at all levels of the company. Every quarter, diversity leaders meet with business leaders to review diversity and inclusion goals, as well as to share current workforce metrics that reflect employment practices in their organization. The data provides demographic analysis regarding new hires, promotion, retention, and representation at various levels. Insights gleaned from the data highlight both strengths and opportunities—in an objective fashion. By reviewing this data with the diversity team, business leaders can make well-informed decisions to help meet their diversity and inclusion goals.

Noted Fan, “The use of predictive analytics equips us with a futuristic view. We need all kinds of talent to sustain our business growth. Walmart took 50 years to get where it is today—a nearly half-a-trillion-dollar business. If we continue to grow at our current rate we will need only 15 more years to become a $1 trillion business. If we want to support that size of business, we need to attract the top talent across the world. Predictive analytics project our future talent needs at various levels, highlighting potential skill and competency gaps, helping us to engineer a long-term talent strategy to support our tremendous growth.”

Walmart does face some challenges in tapping big data. Capturing and mining data across the globe is challenging, thanks to varying local laws and government regulations, differing category definitions, infrastructure limitations, and other concerns. Walmart continues to eye ways to capture data effectively around the globe.19

**International Paper: Narrowing the Scope**

International Paper has long worked to capture talent data to support various goals. Over the years, International Paper has utilized assessment tools based on job success competency profiles and employee development programs to help identify high potential
talent. Third-party resources have been used to identify strong candidates for expatriate assignments, assessing mobility, global acumen, family dynamics, cultural agility, and other factors to ensure a candidate could make a global move.

Recently, International Paper implemented a global workforce-planning tool that helps to calculate retirement risk over a five-year period. The tool uses actual retirement history, age analysis, and straightforward government retirement-age requirements to anticipate retirement trends. For International Paper, having good retirement data is a particularly relevant concern.

“We are facing a huge retiree trend bubble,” said Quinn Thompson, International Paper’s director of global diversity and talent acquisition. “That was a primary factor for us in implementing the strategic workforce planning tool.”

The workforce planning system includes more than 100 analytic tools, ranging from attrition to headcount trending to performance metrics and engagements. Additionally, the system can model various headcount scenarios, incorporating external market data around labor pools, compensation, and related metrics.

Thanks to data captured from predictive analytics, Thompson notes the company has embarked on several recruiting campaigns, increased its employment brand exposure, and launched several social media programs, with an eye toward building and retaining best talent.

“Predictive analytics has helped narrow the scope of what we needed to do and how we needed to do it,” said Thompson. “For example, it narrows down places where we will need talent, and also helps clarify whether available candidates are there, and what compensation in that market is like.”

International Paper’s predictive analytics goals don’t stop with its current initiatives. Said Thompson, “One of our goals would be to expand the research and real-time updating around expat compatibility. I’d also like to increase our understanding of work environment preferences for various groups, such as gender, race, multinational, dual-career, age-based and so on.”

In considering even broader goals and metrics that might apply to all businesses, not just International Paper, Thompson noted that current healthcare costs are a huge focus. “If we can merge the process of educating workers about their personal health management with data such as family history triggers, preventive care practices and overall health and exercise programs, it would be a great way to reduce cost and improve overall quality of life.”

**Big Questions About Big Data in Human Capital Management**

**How to Get the Data Out of Various Systems?**

Too often, people data does not reside in one tool that supports easy aggregation. More frequently, employers house workforce data on several systems. Healthcare data may reside

---

**ANALYTICS LORE**

**RBIs versus OPD**

It’s not so surprising that workforce data increasingly is taking center stage. Some of the dazzling success around Big Data and predictive analytics is directly tied to talent management—in sports.

If you’ve seen the movie *Moneyball*, you know the drill. The movie is based on the real-life story of the 2002 Oakland Athletics, whose managers fielded one of baseball’s most competitive teams that year while paying the lowest payroll. The nerdy, number-crunching character in the movie is, in real life, Paul DePodesta. A one-time talent scout, DePodesta tracked player data across the major leagues, assembling that data in a computer so he could generate reports. He pioneered the notion of assigning value to every action a player made on the field—versus the heavy emphasis on runs batted in (RBIs). Particularly, DePodesta showed that on-base percentage (OBP) was a better predictor of a player’s success than RBIs. The key to DePodesta’s—and the A’s—success wasn’t in shifting from RBIs to OBP in and of itself. Rather, the A’s success stemmed from DePodesta’s ability to capture a variety of performance data, and use it to identify under-valued players who could be an asset to the team.

**Nate Silver: Data Whiz Kid and Presidential Prognosticator**

Another famed statistician with roots in baseball is Nate Silver. Most people today know Silver thanks to his successful call of the 2012 presidential election, where he predicted every state correctly. In 2008 his results were similar, if slightly less perfect: He missed only Indiana, a state in which the margin of victory was razor thin.

In 2012, Silver often was a naysayer, denouncing polls that regularly predicted declines and loss for President Obama. In the end, as noted, Silver prevailed resoundingly, proving the soundness of his approach.

For Silver, like DePodesta, it’s not about the data per se. Nor is it about any one particular measure or performance indicator. Silver’s success came from studying the methodologies used by other pollsters, evaluating their models—noting what data was artificially weighted to offset other outliers—and honing in on successful data strategies. Silver did not conduct polls of his own, but used his own models to aggregate data from other polls, evening out inaccuracies or flaws in pollsters’ methodologies. In other words, Silver’s success, even today, comes not from having data, or possessing a superior source of data, but from knowing how to most accurately crunch the data.

Silver’s history with data is long, and, like DePodesta’s, is rooted in baseball. In 1984, the 6-year-old Michigan resident reveled in the Detroit Tigers’ World Series victory, and found the perfect union of his two interests, math and baseball. After college, while working for a consulting firm, Silver began dabbling in baseball statistics on the side, developing his PECOTA system for predicting a player’s performance and career trajectory. PECOTA (Player Empirical Comparison and Optimization Test Algorithm), designed for both fantasy baseball enthusiasts as well as baseball professionals needing to evaluate players, projects hitting, and pitching performance for baseball players. Rather than honing in on new and little-used statistics, as DePodesta did, Silver’s PECOTA system matches current player data to comparable values generated by previous players—in other words, he is using other players’ past performance to predict the success of similarly skilled players.

Silver has, as well, unique insights into the inclusion challenge. Given his fascination with math at a very young age, he quickly was labeled a geek. On top of that, Silver is openly gay. Asked which made him feel more of an outsider, he replied, “Probably the numbers stuff since I had that from when I was six.” Asked if he believed people view him as a gay icon, he responded, “I’ve started to notice it a little bit, although so far it seems like I’m more a subject of geek affection than gay affection.” He further added in *The Observer*, “I’ve always felt like something of an outsider. I’ve always had friends, but I’ve always come from an outside point of view. I think that’s important. If you grow up gay, or in a household that’s agnostic, when most people are religious, then from the get-go, you are saying that there are things that the majority of society believes that I don’t believe.”
in a repository that is not linked to performance data, which may be housed separately from basic demographic data. And then, where do specific elements of data reside? Would data regarding a person's promotions and career trajectory more likely reside in the performance system or the demographics system?

Is HR Acting as Big Brother?
Yet another dimension is managing data—and findings—appropriately. How much does any employer want to come across as Big Brother, looking over the shoulder and monitoring its workers every move? When workers discover their behaviors are being tracked and, to some extent scored, what happens? What happens when a worker disagrees with his or her score? Anticipating and preparing for these concerns up front can make a considerable difference. According to Laszlo Bock, senior vice president of people operations at Google, “when we look at any data related to our people, we treat the data with great respect. Typically, we give people an option to participate in anything either confidentially or anonymously. The lesson for anyone looking at this space is that you need to construct this really powerful tent of trust in the people gathering the data and how they use it.”

Obligation to Act?
Some companies avoid collecting data because, once collected, action is obligatory. Thus, enterprises that are voluntarily working to increase representation of diverse populations may choose not to actually track progress. What happens when a company begins tracking the progress of women in leadership—and finds it is lagging much more than anticipated? Is the company now obligated to report that? Is the company required to resolve the situation, doing so more quickly than originally planned? As Big Data and its ability to predict success takes root, these worries may begin to fade. In the interim, diversity leaders must be vocal about the need to capture at least voluntary data, whenever possible, related to key populations.

Conclusion
What can diversity practitioners do, faced with the potential for big data, albeit surrounded by myriad questions and possible complications?

The bottom line is people data already exists, in some form or shape, within your organization. And, increasingly, businesses employ data analysts to make sense of that data. Your role should be to help your organization understand the connection between Big Data, predictive analytics, and diversity and inclusion.

If your HR department has moved toward capturing people data and making predictions, it’s time for you to get your own foot in the door. You don’t have to be mathematically astute or even fluent in analytics, although a class in statistics might be helpful. In fact, as a diversity practitioner, it may be better that you’re not, because as the well-known statistician George Box noted in a 2013 Harvard Business Review article, “All models are wrong, but some are useful.” Which means that by themselves, predictive models and algorithms can only try to simplify our complex world—its people, organizations, and processes. It’s your knowledge of the big picture of the human resource function, diversity and inclusion, and the organization that provides balance to Big Data. Ask tough questions about model assumptions, analytical approaches, or alternate predictions. Then, identify some of your key initiatives, or search through your diversity project wish list. Can predictive data support any of your efforts? Can it help you build your business case? Can it help you refine or better focus an effort to deliver more substantial results?

If you’ve answered yes to any of these questions, you need to find ways to begin using predictive analytics. And if your enterprise has not started pursuing data from a talent management standpoint, it’s up to you to change that. Find the places in your organization that do use data—marketing, for example. Get to know the analytics team. Do some investigating, and identify where people data resides in your organization. And then, connect with the analytics folks to see if you can team together to devise some compelling, money-saving ways to tap people data to make your business more successful—and your diversity and inclusion initiatives more successful, too.

ENDNOTES

4 MappyHealth.com website. Link: http://mappyhealth.com/
8 Gardner, Nora et al., “Question for your HR chief: Are we using our people data to create value?” McKinsey Quarterly, March 2011. Link: http://www.mckinsey.com/insights/organization/question_for_your_hr_chief_are_we_using_our_people_data_to_create_value
12 Gardner, Nora et al., “Question for your HR chief: Are we using our people data to create value?” McKinsey Quarterly, March 2011. Link: http://www.mckinsey.com/insights/organization/question_for_your_hr_chief_are_we_using_our_people_data_to_create_value
18 E-mail correspondence and telephone conversations between Susan Welch and Ingrid Jacobs.
29 Gadwaillad, Carol, “Nate Silver: It’s the numbers, stupid,” The Observer, November 17, 2012
30 Grant, Japhy, “Nate Silver on African Americans, Prop 8, and Being A Hero to Queers and Geeks Everywhere,” Queerly, November 21, 2008.
31 Gadwaillad, Carol, “Nate Silver: It’s the numbers, stupid,” The Observer, November 17, 2012
Diversity Best Practices

Diversity Best Practices, a division of Working Mother Media, is the preeminent membership organization for diversity thought leaders to share best practices and develop innovative solutions for culture change. Through research, benchmarking, publications and events, Diversity Best Practices provides members with valuable resources, information, and strategies on how to create, implement, grow, and measure first-in-class diversity programs.

Diversity Best Practices’ services help companies clarify opportunities and implications of their current diversity strategy, identify and enhance critical diversity leadership competencies, create and implement a system-wide focus on diversity and inclusion, and gain the executive-level support needed to ensure the company is successful.

Diversity Best Practices’ team includes an impressive group of relationship managers, researchers, senior practitioners, consultants, council members, and committees from a wide range of cultural backgrounds and professional experience. Our research-based benchmarking content builds the knowledge and offers the tools needed to provide diversity solutions that meet the unique needs of our member companies.

In today’s information-driven economy, diversity leaders need access to the most relevant knowledge available to execute successful diversity initiatives. Diversity Best Practices provides that knowledge. Become a part of the Diversity Best Practices network today. Visit diversitybestpractices.com for more information.