

# MARKLE

## Big Data, Data Analytics, and America's Economic Future

By Dan Wagner, Civis Analytics<sup>1</sup> | July 2013

---

*This paper was prepared for the Markle Economic Future Initiative.*

---

Imagine a man who has worked as a welder for twenty years in Northeast Ohio.

One day, after work, he is unexpectedly laid off from his job. He goes home that day to tell his wife the news. They sit their two children down to tell them that they will have to start making big changes in how the family spends money. The next morning, he begins searching online for welding jobs in Northeast Ohio. For weeks, he sends his resume to generic human resources email inboxes. It takes him days to find any available opportunity that might work, and weeks to hear back, if he ever does at all. Meanwhile, the family's savings dwindle and they wonder what options they have left.

This is a scenario playing out across the country, every single day.

Now, imagine an alternative scenario. The morning after being laid off, when he wakes up, an email is waiting for him in his inbox. That email contains a list of every job posting in a 100-mile radius that fits his predicted qualifications and experience—in this case, complex TIG welding for multiple alloys. The next day, he receives a call from a recruiter in Pittsburgh who has a job opening and is looking for his unique experience. Later that day, the human resources manager at a factory in Dayton calls with a similar opportunity. He will have to make a tough choice about moving with his family, but the initial despair of unemployment has for the moment been put aside by the potential for opportunity. This is the real story of big data.

What happened in this second scenario is that - behind the scenes—an organization saw that he had been laid off, and then matched him (based on his consent) to a set of predicted 'best-fit jobs' based on a text analysis of his job description. But before that, that organization had built a sizeable historical database to calculate a matching algorithm between qualifications and openings. That algorithm said his specific welding experience would fit—on average—with those types of openings. That organization was then sending targeted emails of candidates to recruiters and human resource departments, who would have been humanly incapable of making the same connection in real-time.

For many years—problems like the one above and so many others have remained present, brutal, but unobservable. In some cases we knew they were present, but we had no means to adequately or completely measure them—hospitals couldn't truly evaluate patient outcomes based on different procedures, educators couldn't see the connection between teaching methods and student achievement, and the labor market couldn't easily connect people to opportunities. The potential for big data is that we can begin to see problems as they are—transparent relationships between inputs and outcomes—and build programs to mitigate them through

---

<sup>1</sup> Civis Analytics, previously Obama for America

inference and intervention. The potential for the middle class—in terms of education, social mobility, and economic opportunity—are massive, and we're just starting to learn how.

### **Big Data—A primer from the Obama 2012 campaign**

Every day we are generating, capturing, and storing an increasing amount of data to digitally represent the objects and events that surround us. 'Big Data' is the gradual accumulation of huge amounts of that data into centralized databases for evaluation, inference—and, ultimately, action. And every day, 'Big Data' gets even bigger – improving our digital representation and understanding of the world.

The 2012 Obama campaign represented the first-time a political organization had fully leveraged big data to greatly affect an election's outcome. Our organization created large, integrated databases of voter data, built complex statistical models to predict voter attributes and outcomes at the individual-level, and—most importantly—built a metrics-driven organization that was resource-efficient and demonstrably effective in moving large numbers voters to change their minds and/or turn out to vote.

Prior to 2012, political campaigns had classified voters into broad demographic categories - parsing the country into categories like 'soccer moms' and 'NASCAR dads'. In reality, though, these categories were little more than heuristic stereotypes— 'NASCAR dads' included democrats, republicans, rural voters, suburban voters, and many (probably a majority) that probably didn't even watch NASCAR to begin with. Furthermore, there was little proof that these voters were actually 'persuadable.'

In 2012, OFA used propensity modeling to categorize voters based on the likely outcomes we cared about support, turnout and persuadability. A 'support model', for example, ranked every voter on a scale of 1 through 100 representing their likely probability of supporting the president. The persuadability model ranked voters on a scale of 1 to 10 from the least to most persuadable.

With these models, OFA could define GOTV and persuasion targets at the individual-level and prioritize all activities (field, TV, digital) around those specific target voters. Furthermore, OFA could define its success in terms of clear and almost obvious metrics—if we turn out this precise set of voters we judge our activities successful, if we do not we judge our activities to be a failure.

Those metrics defined the organization from the top of the campaign to the bottom. Every organizer—and even volunteer—saw their work through the prism of registering, turning out, or persuading a set of voters within their geographic area, and they could judge their success or failure accordingly. Headquarters, furthermore, could build national program based on the same (but aggregated) targets. It was as if the campaign were a local-level race, but multiplied by 10 thousand times.

## **Big data and the American Middle Class—A Case Study for an Ohio Welder**

The role of big data in our national problems is to take large, seemingly overwhelming national problems, distill them down to a set of measurable challenges, attach metrics to those challenges, and design individual-level implementation. Our role as policy-makers, organizations, and educators is to use those metrics to define the terms of success for our organizations and work.

The challenge of the middle class is one of those seemingly intractable national problems. The American Dream connotes equal opportunity, but U.S. social mobility—the rate at which individuals rise above their income class at their time of birth – is far lower than other developed countries<sup>2</sup>.

This paper's introduction discussed the case of a high-skilled welder laid off by a Northeast Ohio business and a set of big analytics and execution to help him find the best-fit job. As our increasingly competitive, global economy continues to disrupt traditional middle-class jobs, it is important that we consider how we can use big data to create a more efficient and fair labor market that creates better work outcomes for middle-class people.

In the welder's case, the fundamental issues are information liquidity and execution—there are a set of positions for which the welder is qualified, but there are both knowledge, structural, and execution barriers between them. The consequences for the welder and society are severe: he may go unemployed longer and there is high likelihood that he may choose a lower skilled job, lowering total societal productivity and fairness.

The fundamental role of big data in this case is three fold.

First, it can repair the information liquidity problem between individuals and opportunities. The welder has several available opportunities, but both the individual and the prospective employers are unable to provide a quick 'match' between the two sets. The fundamental role of big, data analysis would be to leverage historical employment data (at the individual level) to determine best-fit positions for this specific welder. Available historical data might include self-reported job information, tax data, employment records and more.

Secondly, it substitutes collective information for individual information. Typical labor market decisions, especially hiring, are executed via single individuals physically evaluating text. This process is time-intensive, mistake-laden and dependent upon individuals that may share different perspectives about candidates and qualifications. An automated process that evaluated opportunities per candidate based on the collective information of thousands of records would provide a check against a potentially flawed process and likely yield more accurate assessments.

Thirdly, it speeds time to execution. Big data has the ability to evaluate potential opportunities within the database, but also to provide that information to individuals directly that might need it.

---

<sup>2</sup> Does America Promote Mobility as well as other nations? Pew: 11/01/2011

Imagining the welder's case across thousands of middle-class workers, we can conceptualize the measurable outcomes (and metrics) that we can influence using big data. First, the number of available candidates to available opportunities by sector and job type (are there good jobs for people?). Secondly, the rate at which candidates are picking the right vs. wrong jobs (are people matching right? Are they switching out of sector at healthy or harmful rates?). Thirdly, execution – how long does the process take by sector and job, and what are the drivers? All these defined metrics can be used to help facilitate our jobs as policy-makers, governments, and businesses.

While this case illustrates critical problems in our labor market and the role of big data in solving them, the welder is not alone. There is a patient in an ER with no recorded medical history, a student in Detroit who's only barrier to getting into Berkeley is the application process, a decorated veteran returning home from Afghanistan without clear job prospects. These are all big problems that can be structurally mitigated through the use of big data – we can record and predict medical conditions for patients, structure effective college application routes for students, and prepare our veterans for their homecoming with recommendations on where they can be successful. These are the stories of big data.