

DATA

You Don't Need Big Data — You Need the Right Data

by Maxwell Wessel
NOVEMBER 03, 2016



Marion Barraud for HBR

The term “big data” is ubiquitous. With exabytes of information flowing across broadband pipes, companies compete to claim the biggest, most audacious data sets. And businesses of all varieties — old and new, industrial and digital, big and small — are getting into the game.

Masses of social, weather, and government data are being leveraged to [predict supply chain outages](#). Enormous amounts of user data are being [harnessed at scale to identify individuals](#) among a sea of website clicks. And companies are even starting to leverage huge [quantities of text exchanges](#) to build algorithms capable of having conversations with customers.

But the reality is that our relentless focus on the importance of big data is often misleading. Yes, in some situations, deriving value from data requires having an immense amount of that data. But the key for innovators across industries is that the size of the data isn't the most critical factor — having the right data is.

It's Not About Big or Small

Uber is often referred to as a big-data success story. There is no doubt that Uber captures a wealth of information. Using the applications it has running in both its drivers' cars and its users' pockets, it has mapped the real-time logistics flows of human transportation.

But Uber's success isn't a function of the big data it collects. That big data has enabled the company to enter new markets and fulfill new jobs in the lives of its customers. Uber's success results from something very different: the small, *right* data it needed to do something very simple — dispatch cars.

In an era before we could summon a vehicle with the push of a button on our smartphones, humans required a thing called taxis. Taxis, while largely unconnected to the internet or any form of formal computer infrastructure, were actually the big data players in rider identification. Why? The taxi system required a network of eyeballs moving around the city scanning for human-shaped figures with their arms outstretched. While it wasn't Intel and Hewlett-Packard infrastructure crunching the data, the amount of information processed to get the job done was massive. The fact that the computation happened inside of human brains doesn't change the quantity of data captured and analyzed.

Uber's elegant solution was to stop running a biological anomaly detection algorithm on visual data — and just ask for the right data to get the job done. Who in the city needs a ride and where are they? That critical piece of information let the likes of Uber, Lyft, and Didi Chuxing revolutionize an industry.

Getting to the Right Data for the Job

Sometimes the right data is big. Sometimes the right data is small. But for innovators the key is figuring out what those critical pieces of data are that drive competitive position. Those will be the pieces of right data that you should seek out fervently. To get there, I'd suggest asking the following three questions as a process for drilling down to the right data.

Question 1: What decisions drive waste in your business? Most businesses have large sources of waste. Consider the world of floral retailing. The average retail florist can sustain spoilage rates of

more than 50% of their inventory. More than *half* of their flowers simply become refuse. So for innovators like UrbanStems and the Bouqs, the data that makes their businesses so disruptive is the data that enables them to eliminate that spoilage. (Disclosure: I invested in UrbanStems.)

In the words of the Harvard Business School's Ben Edelman, "waste makes for opportunity." Whether it's in industrial production, retailing, or legal investigations, figuring out your sources of wasted effort and resources should guide the way toward the right data. Whether it's as simple as identifying predictions you know you make (how much inventory to stock) or whether it requires you to think about the decisions implicit in your business model (how a cab drives around the city at 10 PM), charting out the decisions will point you toward sources of waste.

Question 2: Which decisions could you automate to reduce waste? Once you have your decisions, the hypothetical becomes what you can actually change. Humans are wonderful at making certain types of decisions. When it comes to deciding which campaigns will elicit the most irrational reactions of other humans to branding and marketing materials, humans can be brilliant. These types of decisions should stay (for now) in the hands of people.

But when it comes to making simple, repetitive, operational decisions (like where to send a cab, how to price a product, or how many flowers to order to a floral shop), machines tend to be much better than people. And although many business models of the 20th century are predicated on human control of these decisions, today we can identify the data to automate more of these decisions than you'd imagine.

Amazon, for instance, is rumored to have eliminated almost all of its pricing team, pushing most pricing decisions toward algorithmic control. For most retailers this would be blasphemous. But if Amazon's algorithm works, it would translate to far less spent on discounts, far less inventory piling up in warehouses, and better predictability of new product introductions — each of which would yield enormous competitive advantage.

Question 3: What data would you need to do so? Once you have an understanding of the waste in your legacy system and you've charted out the decisions that result in that waste, the last step is asking a simple question. If you could have any piece of information, however unbelievable, to make the perfect decision, what would it be?

In Uber's case, it needed to know exactly where all the potential riders in the city were in order to automate decisions surrounding where to send drivers and reduce the waste associated with human drivers searching for the next fare. In the case of General Electric's Predix Industrial Internet software, the company aspires to know exactly when a machine is going to break down, helping to automate decisions about maintenance visits and reduce the waste from unplanned downtime. For health insurers seeking to cut costs, they'd love to know the moment that a diabetes patient's blood sugar dips dangerously low, helping to automate decisions around patient interventions and reduce waste surrounding disease mismanagement.

Those are the right pieces of data to seek out. If you arrive at them by crunching a mass of information, that's wonderful. If you arrive at them by building a new app to sense them directly, even better.

Most companies spend too much time at the altar of big data. And not nearly enough time thinking about what the right data is to seek out.

Maxwell Wessel is the general manager of SAP.io, a business unit pursuing disruptive growth at software giant SAP.
