

The Problem with Big Data Isn't the Data



In today's data-driven world, the ability to harness and analyze vast amounts of information has become a critical factor for business success. Big data, characterized by its volume, velocity, and variety, offers unprecedented opportunities for organizations to gain insights, improve decision-making, and drive innovation. However, the journey to effectively leverage big data is fraught with challenges, including data quality issues, infrastructure limitations, and the need for organizational change. This paper explores the dual nature of big data, highlighting its potential while exploring one of the key, but often overlooked, obstacles that must be overcome to realize its full value. By examining the human factors involved in data management and the importance of redefining the relationship between people and data, we will provide a glimpse into how businesses can start navigating one of the underlying challenges of big data.

The Data Paradox



Data.... That simple word generates many different connotations depending on who you ask. For some, data is the key to unlocking the real potential of the digital world. For others, it represents the ability to answer complex questions and address problems and opportunities that we couldn't before. Yet for those trying to figure out how to move their business into the digital future, it represents the nightmare of dealing with decades worth of questionable data practices and bad infrastructure, all while under the gun of delivering on the great promise of data being advertised everywhere. So data is both the answer to our prayers and at the same time the stuff of nightmares, or at least more than a few sleepless nights.

By this point, there are few, if any people, that do not understand the potential in big data to make a step change in business performance, and specifically the ability to apply advanced analytics to large data sets to generate actionable insights or solve complex problems. So, the obvious question is "why aren't we there yet?" For many the answer will lie in the myriad of projects and programs designed to centralize and manage data. However, if the challenge were just in fixing the infrastructure it would have already been solved for most organizations and well on the way to being solved for everyone else. Coming back to a paraphrase of the question of every child on a long road trip "why aren't we there yet?"; the answer is that the problem with data lies with people and not infrastructure.

The Problem with Big Data Isn't the Data

Garbage In, Garbage Out



Before we go any further, let's clarify that statement a bit. Decades ago a phrase was coined that haunts us to this day; "garbage in, garbage out." This phrase, originally used in the late 1950s to describe the challenges with early computers, is at the heart of the real data problem. Simply put, much of the data we have is of questionable

quality, inconsistent, or just missing; and all of this comes down to people. More specifically its directly related to the rigor and diligence with which people collect data, manage, and/or maintain data streams, or how purposeful they are in adding instrumentation for new data streams. But then if everyone understands the value in data, why are we still having a data problem? The simple answer comes down to how people have been conditioned to value data in practice.

In reality, we got ahead of ourselves in the data race. As an industry, we saw the potential in having more and more data and as capturing data became easier, that is just what we did. We started capturing data and lots of it. We added instrumentation to our processing units and equipment, we asked operators to collect more and more data while in the field, and we started viewing reports as data streams. In essence we buried ourselves in data. On the surface, this still sounds great, but here is the problem with all of this data. It takes real work to capture, maintain, and get value from it. As we started tracking more and more data streams, we left employees asking "Is the effort worth it? Is anyone actually doing anything with all of this data?" For years and years, the answer to these questions was often "no". We were collecting data because we liked the idea of having it and what we could do with it; while in reality we often didn't have time to analyze these data streams, much less do anything meaningful with the data. As a result, we started conditioning people to be very selective about which data streams they really worked at maintaining or capturing. This organizational conditioning is at the heart of our data problem today and is the reason that while everyone agrees on the potential value of data, most are skeptical on the practicalities of realizing that value.

The Problem with Big Data Isn't the Data



Framing the Answer to the Data Question

Now that we've gotten to the heart of the real problem with data, let's look at how we start addressing it so that we can get to the pot of gold at the end of the data rainbow. As you have likely guessed by now, the answer isn't creating data lakes or data islands or investing in ingestion software. While these are all important components of the solution, the answer lies with people. More specifically, the answer lies with how the people responsible for collecting, maintaining, and managing streams of data value that data, remembering that we have spent decades training them that much of the data has no real value outside of it making us feel better to have it or worse as a way of policing the activities of our workforce. Therefore, this is where we start; with the relationship between people, the data they are responsible for, and the value they derive from that data.



As we start looking into this more closely, we recognize the specific answer will often be different for every organization but will always come back to generating value in specific data sets for the people charged with managing that data. At times this may actually require a step backwards in terms of the quantity of data being captured in favor of creating space to allow value adding insights from key data streams to be generated. Then as we start redefining our relationship with data from one of a data black hole to one where the data we collect and maintain provides real, demonstrable value to us, we can start adding back data points or creating new data streams along with the mechanisms to process and extract value from these streams.

Leveraging AI to Change Our Relationship with Data



While there are many paths to redefining this relationship between people and data, one that is often overlooked is through artificial intelligence (AI). The simple reason for this oversight, is that AI shines when it has lots of data to work with and therefore people assume that there has to be lots of data for AI to add value. However, this isn't an entirely accurate perception. AI and machine learning can, in some instances, generate valuable insights from smaller, more limited data sets.

The Problem with Big Data Isn't the Data



Now to be clear, AI still thrives when it has abundant, high-quality data to work with, but there is still potentially value to be attained when working with lean data sets (limited in time span, number of data points, or variables being tracked). There will always be a question of sufficiency, because there is a threshold below which AI simply won't work; but once you've crossed that threshold there is value to be obtained, even if the value is simply directional guidance.

For many, the more limited value creation of AI in lean data applications will keep them from going down that path or will steer them towards a more traditional technology (e.g., Excel, First Principles Models, Descriptive Analytics) for analyzing their data. However, this is potentially a short-sighted perspective. One of the values of AI and machine learning, is its ability to grow and improve as more data is incorporated, essentially its ability to evolve. This evolutionary capability is important because it supports the underlying change management requirements of the people responsible for all of that data. More specifically, applying AI to lean data sets for what may initially be only directional insights or for problems that could easily be solved with other technologies, allows people to start getting value from their data and more importantly immediately rewards them as they capture and integrate new data streams into the AI. Alternatively, applying AI to support complex troubleshooting activities where data quality issues can more easily be identified and addressed can provide a similarly valuable organizational starting point.

Wrapping Up

Coming all the way back around to where we started, the problem with data isn't the data or data infrastructure, it's our personal relationships with the data and the value we gain from it. By remembering that our data problem really is an organizational change problem, we place ourselves in a position where we can better start managing the journey towards big data and all of the promised value it holds. Whether you have lots of high quality data going back years or decades or are lean on data and have just enough for some basic insights, it's important to remember that this is an organizational journey and that to be successful, your first steps should reinforce the value to the individuals within the organization for going with you on that journey. As they see value for themselves, they'll in turn become better custodians of the data and thereby start delivering on that promise of business transformation from big data that people keep talking about.



The Problem with Big Data Isn't the Data

About the Author

Kevin Smith



Kevin is the Founder and President of KES Global Advisors, LLC, and has spent more than 25 years driving business transformations for both his clients and with his own teams. His industry experience spans the globe, with direct work experience in over 40 countries across six continents.

Kevin can be reached at ksmith@kesglobaladvisors.com.