How Big Data Liberates Research

There is a tidal wave of conversation about big data. The conversations range from simply defining what big data means, to the business applications of big data, to the societal implications of living in a big data environment. A quick Google search on "big data" provided 1.66 billion results, and I'm sure that number has increased since I wrote this Point of View.

William C. Pink
Senior Partner, Creative Analytics
bill.pink@millwardbrown.com

Out of this cacophony of commentary, one of the most hotly contested topics in our industry is whether big data will replace traditional market research and perhaps make primary research obsolete. It's not as crazy a question as it sounds.

BEFORE BIG DATA

Let me begin with a story. A few years ago, I ran a training class for mid-level researchers about innovations in research. I began by asking the group, "Have you ever responded to an RFP or client request for brand or communication insights by recommending something other than a new survey?" The participants looked at me, puzzled. Finally one of the attendees said, "But we use surveys to measure brand health and communications impact. For every new situation we need a new set of data. How could we *not* recommend a survey?"

Will big data make primary research obsolete?

This was the usual thought process in what could be called the "before big data" world. Whether the research objective was to segment consumer needs to improve targeting or to evaluate the impact of advertising on brand health, it was reasonable to assume that we would need to generate and analyze a new set of data on each occasion. But that assumption is no longer valid. In today's big data world, nearly everything is passively observed and managed

in a digitized fashion; thus we have the ability to use data assets that were previously untapped or nonexistent to quickly and deeply address these same topics.

Big Data isn't really a brand-new phenomenon; for years now, large data sources have included information on customer purchases, credit scores, and lifestyle information. And for years, data scientists have used this data to help businesses evaluate risk and anticipate customer needs. The difference today is twofold: more

In today's big data world, nearly everything is passively observed and managed in a digitized fashion.

sophisticated tools and methods are available to analyze and combine various datasets, and these analytic tools are now augmented by an avalanche of new data sources ignited by the digitization of nearly all data collection and measurement.

The range of content now available is both inspiring and intimidating to researchers raised in the structured survey environment. Consumer sentiment is captured on websites and the variety of social media outlets. Exposure to advertising is recorded not only by set-top boxes but also by digital tags and mobile devices communicating with TVs.

MillwardBrown

SHARE 1

Behavioral outcomes such as call volume, shopping patterns, and purchases are now available in real time. Thus many of the insights that were previously provided by survey research can now be discerned through big data sources. And all of these data assets are generated on an ongoing basis, independent of any research process. These are the changes that motivate the question of whether big data will replace market research.

IT'S NOT ABOUT DATA—IT'S ABOUT QUESTIONS AND ANSWERS

Before we sound the death knell for survey research, we should remind ourselves that it's not the existence of any particular data asset that ultimately matters. What matters is our ability to answer questions. And the amazing thing about the big data world is that the findings from our new data assets generate more questions, and those questions tend to be best addressed by traditional survey research. In this way, as big data increases, we see parallel growth in the presence and need for "small data" to explore and answer the questions it raises.

Before we sound the death knell for survey research, we should remind ourselves that it's not the existence of any particular data asset that ultimately matters.

As big data increases, we see a parallel growth in the need for "small data" to answer the questions it raises.

Consider a setting in which a large advertiser has constant, real-time monitoring of store traffic and sales volume. Existing research designs, in which we probe survey panelists on their purchase motivations and point-of-sale behaviors, help us better target certain shopper segments. Those designs can be expanded to pull in a wider range of big data assets, to the point that big data is the passive monitor and surveys become the focused, ongoing probes into changes or events that require exploration. This is how big data will liberate research. Primary research will not have to focus on what is happening—big data will do that. Primary research can focus instead on explaining why we are observing certain trends or deviations from trends. The researcher can think less about generating data and more about analyzing and leveraging it.

At the same time, we see big data allowing us to address one of our biggest problems, that of excessively long surveys. A wealth of research on research demonstrates that bloated survey instruments have negative effects on data quality. While many have recognized this issue for a long time, the default answer has remained "but I need that information for my senior management," and long surveys have continued apace.

In a big data environment, when survey metrics can be provided by passively observed measures, the issue is moot. Again, think of all the surveys with a focus on consumption. If big data assets are providing insights on consumption via passive observation, primary research via surveys will not have to collect this type of information, and we can finally deliver on the vision of shorter surveys instead of simply providing lip service to that goal.

MillwardBrown

BIG DATA NEEDS OUR HELP

Finally, the "big" in big data is just one characteristic of these new data assets. "Big" references the massive size and scale of the data, which, rightfully, should be front and center, as the scope of big data is beyond anything we have worked with before. But other characteristics of these new data streams are also significant: they are often raw in format, unstructured or, at best, partially structured and riddled with uncertainty. A growing area of data management, aptly named "entity analytics," has developed to help manage the noise in big data. This practice is dedicated to parsing through these data sets and figuring out how many observations are of the same individual, which observations are current, and which are useful and complete.

This kind of data cleaning is necessary to remove erroneous data or noise whether dealing with small or big data assets, but this is not enough. We also need to create context around big data assets based upon our prior experience, analytic strength, and category expertise. In fact, many analysts are pointing to the ability to manage the uncertainty inherent in big data as the source of competitive advantage, since it should result in better decision-making.

And this is where primary research is not just liberated by big data, but contributes to the content creation and analysis within big data.

The ability to manage the uncertainty inherent in big data may be a competitive advantage.

The application to social media data of our new meaningfully different framework of brand equity is a prime example. This framework is validated to in-market behaviors, is implemented on a standardized basis, and is easy to extract into other marketing operations and information systems to support decision-making. In other words, our equity framework, powered (though not exclusively) by survey research, has all the properties needed to overcome the unstructured, unconnected, and uncertain nature of the big data.

Primary research is not just liberated by big data; it contributes to the content creation and analysis *within* big data.

Consider data on consumer sentiment provided by social media. In its raw form, the peaks and valleys of consumer sentiment are often minimally correlated with offline metrics of equity and behaviors; there is simply too much noise in the data. But we can reduce that noise by applying our constructs of consumer meaning, differentiation across brands, dynamism, and salience to the raw consumer sentiments as a way to process and aggregate the social media data along these dimensions.

Once the data is organized in alignment with our framework, the resulting trends typically align with offline metrics of equity and behavior. In effect, the social media data could not speak for itself; it required our experience and constructs around understanding brands in order to be

Our equity framework, powered by survey research, has all of the properties needed to overcome the unstructured, unconnected, and uncertain nature of big data.

leveraged to that purpose. Social media brings us full circle when it provides unique findings on the language consumers use to describe brands, and we then bring that language back into our survey designs to make the primary research that much more effective.

THE BENEFITS OF LIBERATED RESEARCH

This brings us back to how big data is not replacing research but rather is liberating it. Researchers are liberated from having to generate a new survey on each new learning occasion; ongoing big data assets can be leveraged for many topics, allowing subsequent primary research to go deeper and fill in the gaps. Researchers are liberated from needing to rely upon bloated surveys and instead can keep surveys short and focused on those variables that they are ideally suited for, resulting in better data quality.

Once liberated, researchers can use their established first principles and insights to impart accuracy and meaning into the big data assets, leading to new areas of survey-based exploration. This cycle should lead to deeper insights across a range of strategic issues, ultimately moving toward what should always be our primary objective—to inform and improve brand and communications decisions.

To read more about big data and market research, please visit www.mb-blog.com.

If you enjoyed "How Big Data Liberates Research," you might also be interested in:

"Social Measurement Depends on Data Quality and Quantity"

"The Big Risk of Big Data"

"Tracking at the Crossroads"

"Beyond the Bias in Big Data"

"Where Will Big Data Make a Big Difference?"

"Why I Believe Big Data Needs a Little Help"

MillwardBrown