

Making Big Data into better data

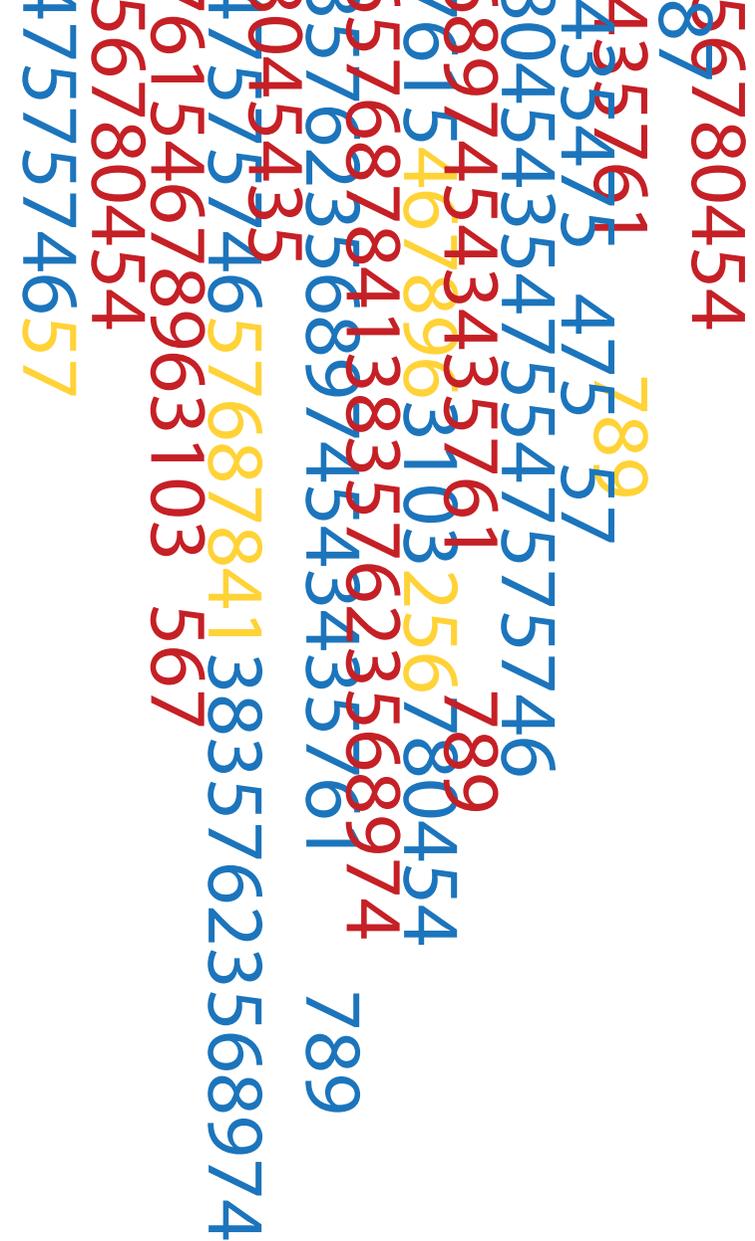


by **Mark Kingsbury**
Global Head of Marketing Science,
TNS



and **Bob Burgoyne**
Development Director, Marketing Science,
TNS

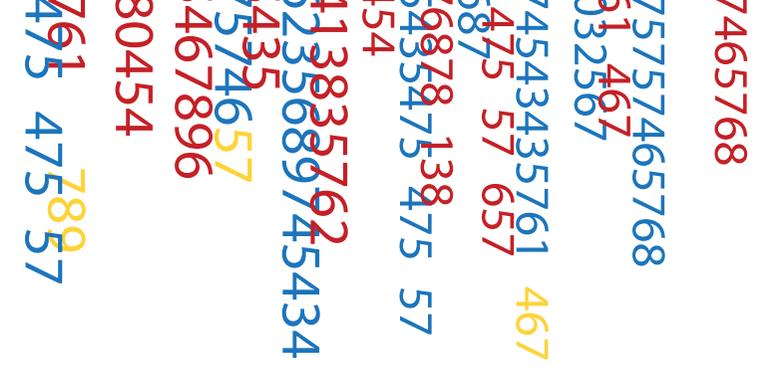
It may be immense, fast and mind-bendingly varied. But to make the most of Big Data we must remember that it cannot speak for itself.



Share this [f](#) [t](#) [in](#)

Intelligence Applied – special edition
Autumn 2013

Making Big Data into better data



Mind and the machine

Big things can be intimidating. Marketers and researchers cannot allow Big Data to be one of them. It's true that the numbers involved are staggering: 90 percent of the data that has ever existed was created in the last two years and 'data taps' such as mobile, social and POS will continue to pour out raw information for researchers to work with at an ever faster rate. It may be tempting to conclude that human intuition must surely give way to computers and algorithms when it comes to keeping up. But if our response to this wave of data is to retreat behind number-crunching technologies then we will have missed a huge opportunity. New data sources have the potential to transform the role of research and expand our understanding of human behaviour. However, they can only do so if we continue to apply the immense, unique power of our own minds.

So what do we mean by 'Big' exactly?

Big Data would feel a lot more manageable if it were just a question of having more numbers to deal with. But Big Data is bigger than that. It represents the coming together of several different themes, each of which would be fairly paradigm-shifting in its own right.

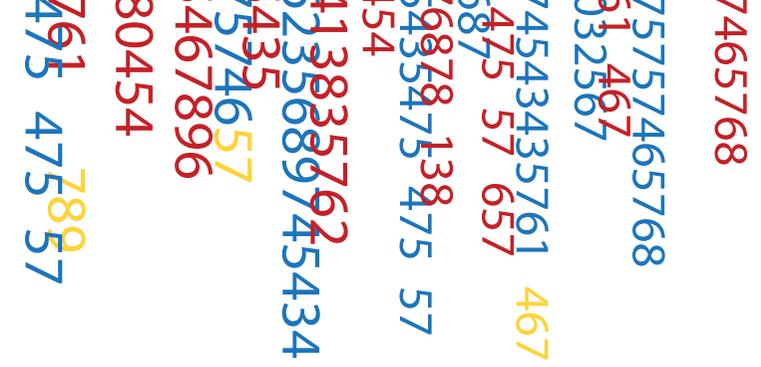
First of course, is the sheer scale of the data now being produced and stored. Walmart currently handles more than 1 million customer transactions every hour, in databases estimated to contain more than 2.5 petabytes. Such an organisation may soon have created more data than research surveys have ever delivered.

Data's velocity, the speed at which huge volumes of it can be generated, is every bit as breathtaking as

its sheer size. Data now generates itself; it is created and stored simply by virtue of things happening. And as a result there are no limits to how big it can get and how fast it comes at us.

Yet perhaps the most challenging shift of all is that this size and speed is combined with an explosion in the variety of data forms. Big Data comes in all shapes and sizes. Researchers are leaping on new sources of data – and new sources of data are leaping on us: from mobile activity to Twitter feeds, geo-location information, facial expression capture and much more. We are quickly moving from dealing in numerical scores to dealing in shapes, movement patterns, expressions – and human language. And such data does not come readily packaged for analysis; using it must involve translating it as well.

Making Big Data into better data



You created it: you deal with it

Faced with such challenges, it's tempting to believe that computational power, which has taken the lead in creating this new world of information, must also take the lead in defining how we deal with it. In this view of the world, the researcher starts to look less like a person, more like supercomputer in a bunker: one where we simply have to feed in the right question or combination of questions, plug it into the river of Big Data – and wait for the answer to pop out. But there are significant dangers to this approach. If Big Data ends up becoming processed and commoditised data, then we are all in trouble.

Digesting really raw data

It's a mistake to believe that data can ever speak for itself. Data always speaks with a human voice; it can't say anything otherwise. Every statistic that we deal with is the result of subjective judgement about the problems that we should try to solve, what we think the answers should look like, and what data forms we

can enlist to help provide those answers. And these judgements are human ones.

Big Data would feel a lot more manageable if it were just a question of having more numbers to deal with. But Big Data is bigger than that.

In the Big Data era, the human imagination continues to play an essential role in envisaging what our many different data sources can be made to do, and in aggregating, translating and coding them to enable them to do it. To take a very simple example, Google can predict a flu epidemic by spotting spikes in searches on cold and flu remedies. This is a tremendously cool thing, but it only works because

somebody realised that this pattern is significant – and that it correlates to something meaningful and useful. Similarly, micro-location data gives TNS a powerful new tool for mapping movement around stores – but it is only powerful because we have established an understanding of what these movements mean.

In his book *The Signal and The Noise*, US election poll guru Nate Silver points out that data does not arrive in ready-made, readable patterns. We must come to it armed with models and ideas based on our existing insights and understanding. Silver devotes a chapter to global warming and the fact that it would be impossible to find any evidence of this in the notoriously unstable climate record, where annual temperatures fluctuate hugely, were scientists not armed with a theory telling them exactly what to look for – and which data to prioritise. It's an important reminder that the bigger data is, the more it needs help to become articulate.

Making Big Data into better data

creators

In the Big Data era, we are data curators, working with information that has been generated independently.

From data creators to data curators

In the Big Data era, we are no longer data creators, designing the structure of information from the outset, through the crafting of questionnaires; instead we are data curators, working with information that has been generated independently. As such, we will face many new challenges and require many new skillsets. However, as we evolve the role of research, the skills that once went into structuring surveys will remain crucially important in aggregating and selecting data sources, and deciding exactly how they relate to one another. For now, this might involve incremental

improvements such as linking spend and retention data to customer experience surveys, as we already do at TNS, or using listening apps on mobile phones to record actual exposure to TV ads rather than relying on recognition metrics.

In the future, we will find more and more scenarios where the data we aggregate does not include traditional surveys at all. When we start to plug observational data directly into look-alike models, for example, we start to change the role of research, making it a key element within media planning and buying.

But in all of these contexts, it's not just a question of being excited about what data can do. It's equally important sometimes to step back, look at how complete and representative a given set of data is, and ask ourselves rigorous questions about what questions it is best qualified to answer.

The continuing evolution of analytics

At TNS, we've already evolved from the era of ad-hoc analysis, when researchers collected data with little reference to how it would eventually be used (and then looked through it in the hope it would reveal something useful). Today the design of the

7465768
75757465768
31467
32507
74543435761 467
475 57 657
6878-138
435475475 57
54
13835762
235089745434
435
574657
467896
30454
491 475897

Making Big Data into better data

instruments for a particular piece of research is informed from the start by the challenge of how best to answer business questions.

The conceptual framework that we use for any type of analysis reflects how the human brain naturally makes sense of information. This framework consists of four different ways of looking at any set of data. 'Dimensions' and 'landscape' address the structure of information; the first seeking out common themes across a data set (the key themes defining a product category, for example), the second looking more closely at competitive relationships, owned and disputed territory and areas of opportunity. We then build on this structural understanding with more action-oriented means of addressing the data: 'groupings' to segment the subject matter and 'drivers' to reveal the variables that influence relevant results, including causal connections that can be far from immediately apparent.

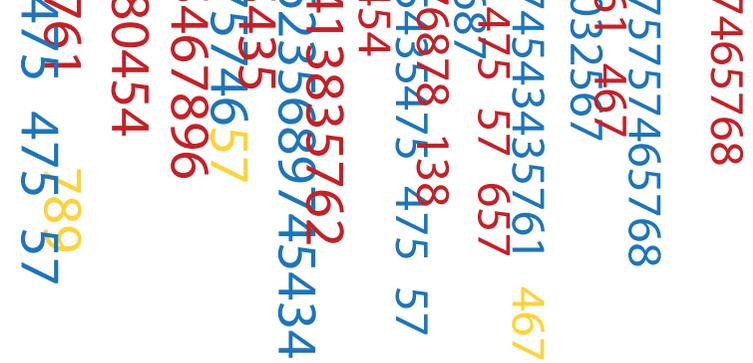
This approach provides a checklist for where and how to look for patterns and themes. In the Big Data era, we will learn to look for different types of patterns in vastly diverse forms of data, but human reason remains the key driving force in identifying them and drawing purposeful connections between them.

Computational muscle can give research the scale and speed that we will increasingly require in the Big Data era, but it is important to distinguish between automating processes and expecting machines to design them in the first place.

We must not fool ourselves that Artificial Intelligence (AI) is ready to take on the task of formulating questions and crafting the algorithms to answer them. After all, even those that welcome the concept of a technological singularity in which

human-designed AI surpasses that of humans themselves, don't envisage it happening until at least 2045. That's a long time to wait to take real advantage of Big Data.

We must not fool ourselves that Artificial Intelligence (AI) is ready to take on the task of formulating questions and crafting the algorithms to answer them. Even those that welcome the concept of a technological singularity in which human-designed AI surpasses that of humans themselves, don't envisage it happening until at least 2045.



Making Big Data into better data

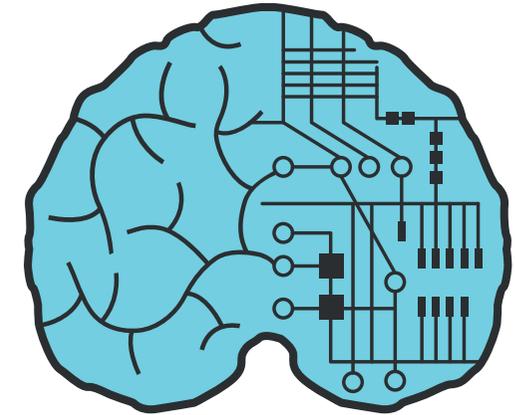
Data and the human imagination

Imposing structure on Big Data will throw up some intriguing challenges – and these challenges will involve logical leaps and lateral thinking for which the human brain remains our best available tool. What is a meaningful way of scoring a positive tweet or Facebook rant? What aspect of somebody's location is actually relevant to the client brief – and what other sources of information can be integrated or overlaid to give context to this information? The location of a car by itself is meaningless. If it's a car unable to fit into the WalMart parking lot on Black Friday, it becomes a whole lot more interesting.

When we talk about deploying computational power in the Big Data era, we must therefore be pretty clear about what we are asking computers to do. Depending too much on non-human processing power risks confusing correlation with causation and failing to distinguish between relationships that are meaningful and those that are not, and it risks setting narrow parameters for our thinking that we forget to look beyond.

We must continue to exercise our judgement as to which information is valid and valuable, and how its many varied forms can be coded in meaningful ways. As data curators, that's our job. But by unleashing the power of today's machines we can dramatically increase the scope of data that we can use, the range of questions that we can ask, and the speed with which we can answer them. When Big Data is aligned properly with human insight and human reason, it can unleash their potential in ways never envisaged before. ■

Read more about our perspective on Big Data [here >](#)



7465768
75757465768
31467
32507
74543435761 467
475 57 657
687
6878-138
435475 475 57
54
13835762
235089745434
435
574657
467896
30454
491 475897