THE BIG DATA PANDEMIC

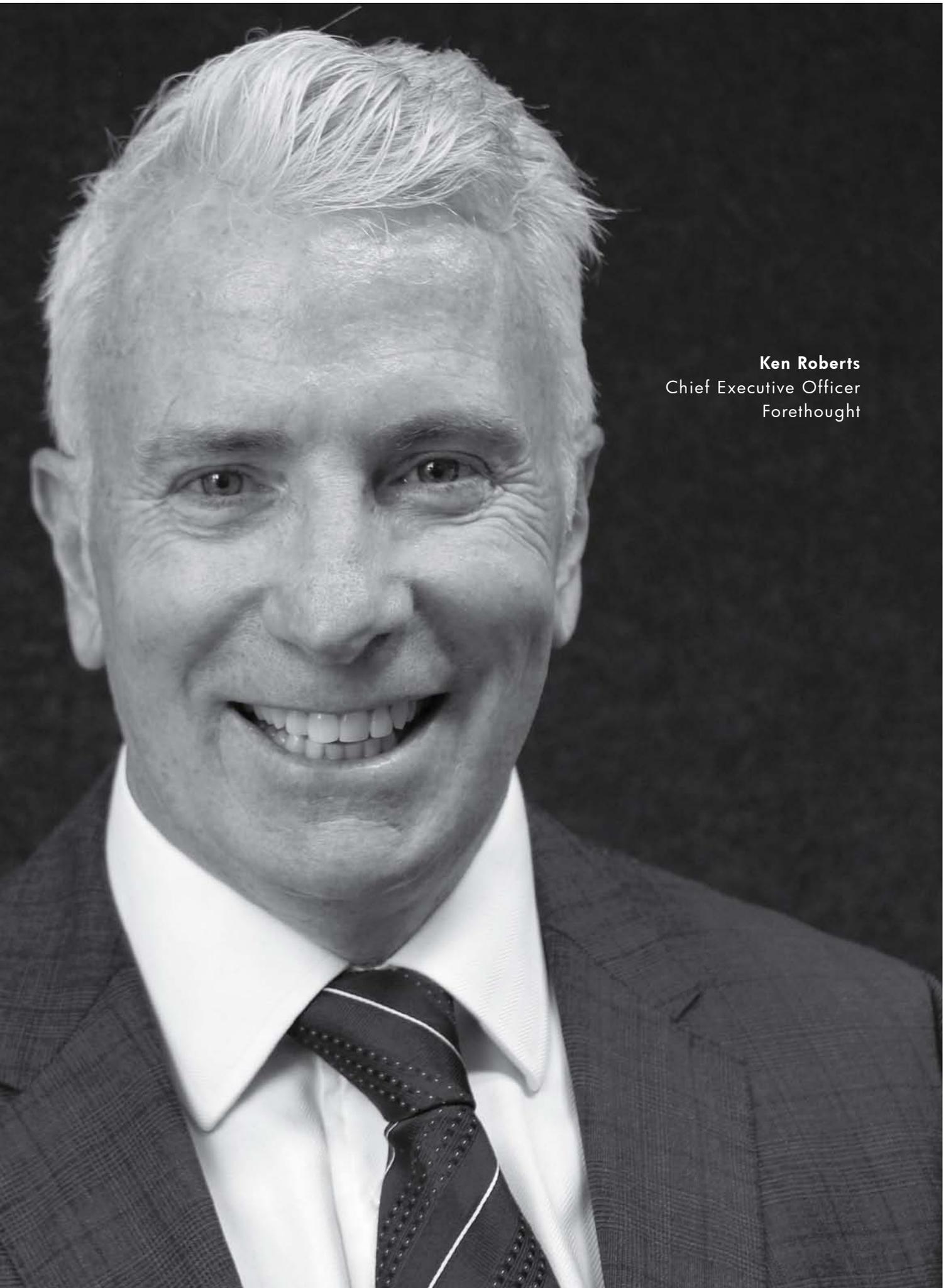## KEN ROBERTS

CHIEF EXECUTIVE OFFICER, FORETHOUGHT

When organisations are
failing to operationalise the insight
from small data, it leads me to wonder
how big data avoids becoming another
management panacea.

forethought®

Ken Roberts
Chief Executive Officer
Forethought

I often wonder if herd instinct is a core subject taught at business schools. Recently, Forethought has been witnessing the adoption velocity of big data as vendors go house to house evangelising the next managerial panacea. Although there has been an abundance of data for a long time, the big data pandemic is only now in full swing.

Big data is not so much defined as the size of the data repository as the complexity of the means for drawing inference from the data. Unlike 'small data'[ii] which has as its stock-in-trade hypotheses, surveys and sampling, big data unleashes largely unaccompanied analytics to discover patterns and correlations. Indeed, some proponents believe that the advent of big data spells the end of surveys and therefore small data.

According to Microsoft researcher and MIT Professor, Kate Crawford, big data is the '*jargon du jour*'. In an HBR Blog[iii], Crawford declared that 2013 looks like the year we will reach 'peak big data hype'.

One would think that investment hurdle rates and payback periods would check the progress of the pandemic; and yet, leaps of faith fuelled by vendors bestowing the virtues of big data continue to penetrate businesses at an increasing rate. Indeed, the big data market is projected to globally reach $18.1 billion in 2013, an annual growth of 61%[iv]. Dr Crawford describes this unchecked faith as 'Big Data Fundamentalism', where there is a belief 'that unless you have really big data sets you are not getting close to objectivity and truth.'

BIG DATA FUNDAMENTALISM

One of the first apostles of the big data fervour to predict the end of small data was Chris Anderson. In his 2008 article entitled 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete[v]', he observed:

> 'This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out the door with every theory of human behaviour, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.'

Apparently Anderson had not considered the importance of what to say to these customers once big data had identified them. That is, to drive consumption behaviour, what should the marketer teach prospective and existing customers and how to make them feel?

Presumably, Anderson would just blast the market with rational offers, and we would see marketers return to the reach and frequency driven gross rating points to maximise awareness and assume all else would follow.

Anderson argues that we can stop looking at marketing models. 'We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where [marketing] science cannot.'

Is he right? Not yet. Examples of errors in big data can be readily sourced. For example, Nature[vi] recently reported that Google Flu Trends had markedly overestimated the proportion of North Americans to get the flu in 2012/13 (See Figure 1). Nature reported that for big data based Google Flu Trends, 'The latest US flu season seems to have confounded its algorithms. Its estimate for the Christmas national peak of flu is almost double the estimate from the USA Center for Disease Control and some of its state data show even larger discrepancies.'
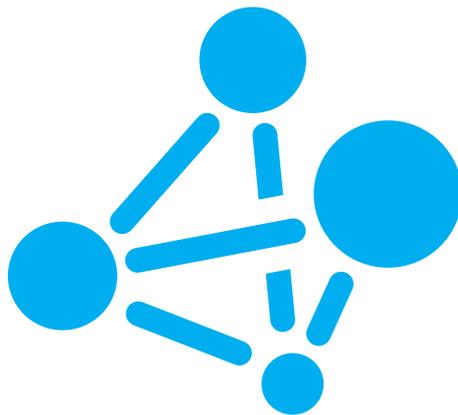
GOOGLE FLU TRENDS

Sources: Google Flu Trends (www.google.org/flutrends);
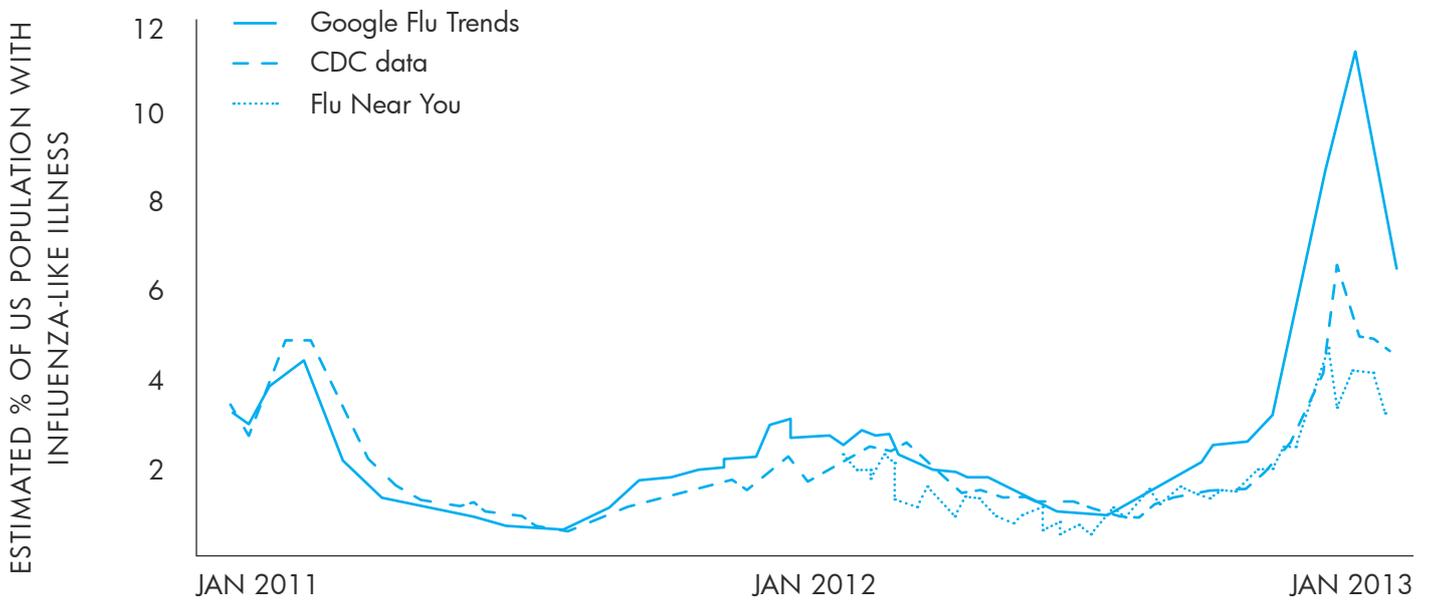CDC; Flu Near You.

Figure 1.
Google Flu Trends.

THE HUMAN FACTOR

Ideally it is the integration of big data, software and small data's great marketing scientists that brings about the best outcome. Too often, in the big data sales literature, the human component to the successful application of big data is understated.

As Jim Stikeleather, Chief Innovation Officer at Dell, has written in a HBR blog[vii], 'Machines don't make the essential and important connections among data and they don't create information. Humans do.' He goes on to say that we should 'understand that expertise is more important than the tool. Otherwise the tool will be used incorrectly and generate nonsense.'

(This raises the question: Are our universities producing graduates with sufficient commercial awareness and mathematical literacy to feed the big data vendors' 'solutions'?)

The value of the interaction between big data and small data analysis has been demonstrated numerous times by Forethought.

For example, on behalf of ING DIRECT Australia, a hierarchical model was produced to highlight the drivers of customer engagement. An issue with a particular service interaction that was highlighted in small data analysis enabled big data analysis to discover that a large proportion of dissatisfied customers had been through common online queries.

AustralianSuper had interrogated its big data for the cause of member defection. What was missing from the big data algorithms was the kind of dependent variable specifically crafted to address the exact situation of defecting members. When the small data survey result was appended to the big data, it was found that seven of the 12 statistically significant explanatory variables were found in the organisation's data repository. Also, five of the explanatory variables were not found in the AustralianSuper data repositories, highlighting benefits of a marriage between big and small data.

Perhaps it is not about big data versus small data but rather, big data and small data combining to produce synergistic insight.

SMALL DATA ACCURACY

If you are not a big data fundamentalist, then perhaps you have applied small data marketing models such as customer engagement to optimise your operational performance, or perhaps you have applied response models that have enabled you to optimise your media mix.

Forethought has dozens of driver level models that have predicted purchase behaviour and therefore changes in market share as reported by third party research vendors with correlations of better than .75. Moreover, Forethought and many other research firms have countless driver level models that use survey data to predict the next brand choice within the survey data with correlations of better than .9.

So if, as the Fundamentalists appear to suggest, small data is so incapable of saying how consumers will behave, how then can multivariate models using survey data so accurately predict behaviour?

Presumably, this strong predictive capability of small data's surveys and analysis is because there is a close to perfect relationship between what we, the consumers, non-consciously decide and what we predict will be our behaviour. Otherwise we would be constantly astonished by our own behaviour.

COPERNICAN REVOLUTION...
YET OLD HABITS DIE HARD

If small data models can be so accurate, then why the need for big data in marketing? The answer is not necessarily found in the poor performance of small data-based behavioural sciences.

The answer may lie with the current generation of marketing management who are not availing themselves of even the traditional capabilities of existing small data analysis.

Rory Sutherland, Executive Creative Director and Vice-Chairman of OgilvyOne London, sums it up:

" The record of the marketing services community to what seems to be a Copernican revolution in the behavioural sciences has so far been mostly notable by its absence. The past reaction to earlier work by Ehrenberg, Jones, Stephen King and so on – which challenges assumptions with real empirical evidence – suggests that marketers may do what they usually do: show great interest and appreciation of this new information before carrying on doing what they have always done.

UK research leader John Kearon provides further illumination; 'The desire to adopt new models and methods may be strong but old habits die hard and the traditional research approaches are hard to relinquish, even when there's an awareness of how bad they are for the company's long-term health.'[ix]

So, if marketing management is not availing itself of small data, how will their adoption and usage be different for big data?

It is for this reason that the failure to fully operationalise the insight from small data leads me to suspect that big data is just yet another panacea.

In marketing, one can't help but notice the arrival of big data. It is not so much the new insight as much as the chunky invoices for big data analytics software and those new, introverted, fresh-faced, back room mathematicians walking the halls. But there are a couple of important questions that need to be thought through.

Will big data outperform small data? The evidence is not in; yet even if it does, how will those insights make their way into the risk–averse management decision and actions?
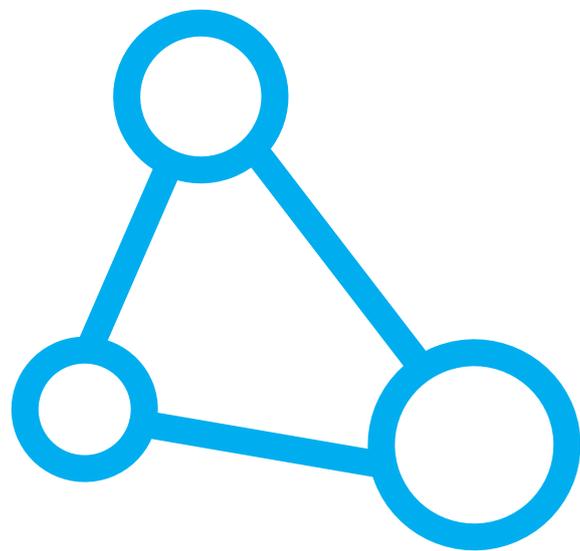
FORETHOUGHT

REFERENCES

i.   Ken Roberts is President and Founder of Forethought - Melbourne, Singapore, New York.

ii.   Small data is defined as marketing modelling based on survey data. For example, a customer engagement model.

iii.   The Hidden Biases in Big Data, April 1, 2013, Crawford, K., http://blogs. hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html

iv.   Big Data Vendor Revenue and Market Forecast 2012-2017, Apr 17, 2013, Kelly, J., http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_ Forecast_2012-2017#Big_Data_Definitions

v.   The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, June 23, 2008, Anderson, C.,http://www.wired.com/science/discoveries/ magazine/16-07/pb_theory

vi.   When Google got flu wrong, 13 February 2013, Declan Butler, Nature, http://www.nature.com/news/when-google-got-flu-wrong-1.12413

vii.   Big Data's Human Component, September 17, 2012, Stikeleather, J., http:// blogs.hbr.org/cs/2012/09/big_datas_human_component.html?utm_ source=pulsenews&utm_medium=referral&utm_campaign=Feed:+harvardbusi ness+(HBR.org)

viii.   Decoded, Barden, P., Wiley, 2013.

ix.   BrainJuicer Group PLC Annual Report and Accounts 2012, March 2013, p.3.

forethought®

FORETHOUGHT.COM.AU