
ACC DOCKET

INFORMED. INDISPENSABLE. IN-HOUSE.

Deluged by Data

Technology, Privacy, and eCommerce



Your average corporate executive has an amazing quantity of data at her fingertips.

There's financial and manufacturing figures, including up-to-the-minute sales information, logistics scorecards, factory orders and inventory tools. She can access website clicks and views, social media heat maps and verbatim customer comments. In some cases she can even see how many consumers are buying her products, by individual store, updated every second.

It's an avalanche of information. No one can read, much less comprehend, all the available data.

But so what? More information is better, right? And there are bright mathematicians out there busily creating algorithms to organize this morass. Technology will save us.

There is, however, a dark side to the information bonanza. The mathematical tools used to sort through these troves of data are exceptional utilities, but they are merely tools. The dark side of big data has nothing to do with the data itself and everything to do with the humans who use it.

Take, for example, the wonderful world of Google Flu Trends.

If you type google.org/flutrends into your browser you will find one of the more innovative examples of corporate information being used for the public good. Google tracks the search queries of users around the world and, with comparisons to historic data, anticipates the rise of influenza cases by region. The theory is that more consumers will search for information about the flu when they start feeling ill, well before the local health authorities are even aware of an epidemic.

For public health professionals, schools and businesses, Google Flu Trends can speed up response

time and preparation. It is a great example of how billions of tiny data points, gathered together, provides useful predictive information.

But naïve reliance on easy answers is risky. In retrospective analysis from 2013 and 2014, the correlation between the prediction of Google Flu Trends and actual influenza rates was surprisingly poor. In some cases the predictions were completely wrong. This could have been from a variety of factors — the type of search terms used, the structure of the algorithm, or even unrelated news about epidemics — but it showed the limitations of mass data assessment.

Google is a responsive and responsible company. They updated the algorithm. The Flu Trends site now presents data that is cross-referenced with traditional public health reports. It remains a useful tool.

But this critique is important for another reason. It reminds us that no matter how much data we collect, no matter how sophisticated our software, we remain human: fallible and easily fooled.

There are distinct benefits to computer analysis of large data sets. Computers are not biased or limited by human prejudices. They do not get tired or make stupid mistakes. But they are used by people that do. Human beings have to write the programs, decide on the data inputs and design the mathematical models. Human judgment brings human error.

The real danger, however, is when normal people like us blindly rely on the results. We like simple answers. We exaggerate the reliability of accessible information. We believe that the headline must be true because it came from someone who is really smart, with a really powerful computer.

And that's our mistake, not the programmers. Google is very careful to say that Flu Trends is not a replacement for real epidemiological data. You cannot tell if someone is sick just by what they type in a browser search bar. But that warning does not stop me or millions of others who live in blind faith and subscribe to technological infallibility.

Which brings us back to that busy corporate executive, awash in data. Like us, she relies on others to interpret all the information around her. As the amount of information increases, so does her dependence on interpretative tools and summaries. And the simpler the answer — especially one that distills complex troves of data — the more she will believe it.

Big data is enormously useful, but to ignore the way humans rely on the “easy answers” of big data can be dangerous.

We shouldn't feel bad, however. Remember that some of the brightest programmers and engineers at Google are working with the largest data sets available, and they make mistakes all the time.

[Bill Mordan](#)



General Counsel and Company Secretary

Shire PLC