



---

## Transaction Banking by D Sign

### *Blog*

Professor Michael Mainelli

December 2014

### **“Big Deeds Need Big Theories For Big Data”**

*As far back as two million years ago primates pursued three basic strategies. Find bright, shiny fruit. Avoid snakes. Introduce yourself to any appropriately attractive partner for potentially intimate interaction. Simple really. Since the information revolution things have gotten complex. In the 1990's a group of financial and technology firms created a project working with Ministry of Defence scientists, the Financial Laboratory Club (F£C). Their objective was to create a simple virtual reality world for financial traders that visualised financial risk.*

*One group held that symbolic abstract worlds are impossible to visualise, noting that we lack simple symbolic abstract conventions for recording broadcast television programmes. That's why people find recording so tough. Another group said such a world was unnecessary as spreadsheets are abstract but realistically represent the situation traders face. A third group wanted to build a fighter pilot or race course simulator where traders would fire at situations they wanted or slow down when the track narrowed. A fourth group wished for some simple guide rules about risk displayed, like road signs, symbolic but analogous to the world traders faced. One of the Ministry of Defence scientists pointed out that the best simulator we could build would appeal directly to the inner primate – “find bright, shiny fruit; avoid snakes; have dates”. If the computer scientists could work through the mathematics and software algorithms to produce that, and he seriously doubted they could, then we wouldn't need the primates in the first place. We settled for some pretty pictures.*

### **Vanquish Your V's**

Many presentations on Big Data are full of v's – “visualise a vast volume and variety of variable data producing value and veracity at velocity despite volatility”. Whew. Or perhaps “vandalise a vapid vacuity and vanity of venal data producing vapour and vexation despite vocals”. For those of us who have followed the trajectory from artificial intelligence, to machine learning, to information analytics, to predictive analytics, to dynamic anomaly and pattern response, to Big Data, it's really not that new, just high-volume in the auditory sense. But might there actually be something larger here?

Big Data is a term for processing data sets so large and complex that new techniques are required. Big Data looks beyond organisations so increasingly awash with data that they are bailing it out of the organisation. Big Data emerged from Big Science issues, the Large Hadron Collider and the Human Genome projects. Big Data looks onward to the seemingly intractable processing problems if scientists and firms could use it all. This is a noble goal and worth pondering.



---

ICT people in science, government, finance and elsewhere should celebrate that Big Data has made it into the management fad folder. Sure, it's frustrating that managers return from their latest conference with "Heard about Big Data? Well, we need some, and fast." It's especially vexing that management have turned down so many sensible ICT investments over the past few years that would have put the solid foundations for Big Data in place. Still, ICT people should celebrate the fact that Big Data finally has management believing that their data has value and may want to do something about it.

But there are two cautions worth noting, the focus on visualisation and management's desire for 'quick wins without pains'.

### **Pretty Pictures Provide Precious Paucity**

A lot of the oohs and ahhs around Big Data follow some wonderful pictures and videos, all the airline flights in the world, sources of attacks on computer networks, the logistics of a large bookseller. They are indeed pretty, but are they useful? I have asked fellow researchers repeatedly at conferences what value the big pretty pictures provide. "Precious little" is the answer, though there is one thing we shall come to in a minute.

They look great, so they must be valuable? Well, it's a bit like showing you a picture of a forest today. Tomorrow I show you another picture of the same forest from the same angle and tell you that 37, 235, 842 leaves have been replaced. What does that tell you? A few months later I point out that many of them have become red or brown. You point out to me that you don't need pretty pictures to work out it is autumn. What I should be pointing out is an unusual die-back among such-and-such a species, or an odd location that is blooming when it shouldn't.

A more commercial example would be the Risk Management Association's key risk indicator (KRI) study, which spun out KRiEX (see [www.kriex.org](http://www.kriex.org)). Some 50 banks at one point defined 1,809 KRIs to be watched every day in every way. So obviously big visualisation would put all the KRIs into some red-amber-green (RAG) display for the board room. Trust me, 1,809 KRIs is also a forest of leaves. Each day the RAG chart changes but you learn little. It's a bit like watching a sculpture of traffic signs blinking (how *a propos* that there is such an installation at Canary Wharf in London - [http://en.wikipedia.org/wiki/Traffic\\_Light\\_tree](http://en.wikipedia.org/wiki/Traffic_Light_tree) & <http://www.wharf.co.uk/2014/01/canary-wharfs-traffic-light-tr.html>).

We need to be directed at specifics within the pretty pictures. We need systems that can examine large data sets and tell us what patterns and anomalies we should examine. This is where dynamic anomaly & pattern response systems (DAPR) excel. DAPR systems would point out that certain risks are occurring together, or that it is unusual to have such a 'red' risk at this time of year. There are many approaches to doing this. Z/Yen has had great success just using support vector machines to provide a basic architecture for doing this regularly, easily, and inexpensively. Adding DAPR systems to Big Data is helpful, but insufficient to unlock the hidden value in data.

### **Big Daddies Needs Big Brains**



---

One of the more interesting debates surrounding Big Data began with a 2008 article by Chris Anderson, “with enough data, the numbers speak for themselves”. [Chris Anderson, “The End of Theory”, Wired (23 June 2008) - [http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory)] He went on in this provocative article to state, “faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete”, “with enough data, the numbers speak for themselves”, and “Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.”

In an excellent rejoinder, Tim Harford took apart four over-simplifications, “that data analysis produces uncannily accurate results; that every single data point can be captured, making old statistical sampling techniques obsolete; that it is passé to fret about what causes what, because statistical correlation tells us what we need to know; and that scientific or statistical models aren’t needed”. [Tim Harford, “Big Data: Are We Making A Big Mistake?”, Financial Times (28 March 2014) - <http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2xWU19KwN>] Anderson had a point that we could let algorithms and machines do a lot of the work. Harford emphasised that “theory-free analysis of mere correlations is inevitably fragile”. Big Data is not the death of science. Geoffrey West, Distinguished Professor and Past President of the Santa Fe Institute, summarises “‘Big data’ without a ‘big theory’ to go with it loses much of its potency and usefulness, potentially generating new unintended consequences.” [Geoffrey West, “Big Data Needs a Big Theory to Go with It”, Scientific American (16 April 2013) - <http://www.scientificamerican.com/article/big-data-needs-big-theory/>]

And it is here that management need to start thinking. The big daddies of the board room need to provide testable hypotheses that Big Data technicians can model and test. A good example of this is a large commodities trader that asked a tough question, "why can't we predict the losses and incidents flowing from today's trading?" The idea was to look at the environmental and activity statistics for each day and use multi-variate statistics to see how strong the correlation was with incidents and losses flowing from that day. “Correlation doesn’t prove causation”, but "correlation should cause questions". This large commodities firm made predictive accuracy the primary measure for its operational risk team. They used operational activities – number of deals, staff hours at desks, telephone traffic – to predict what would happen.

This approach, Environmental Consistency Confidence, says, "if you can predict incidents and losses with some degree of confidence, then you have some ability to manage your risks". You are confident to some degree that outcomes are consistent with your environment and your activities. The converse, "if you can't predict your incidents and losses", implies either that things are completely random – thus there is no need for management – or that you're collecting the wrong data. Knowing that incidents and losses are predictable leads to application of the scientific paradigm. From a proven hypothesis, financial risk tools such as culture change, controls, training, process re-engineering or risk costing can be usefully applied.

### **Big Gain Needs Big Pain**

What’s needed now are Big Ambitions, that result in Big Theories that can be tested. A few thoughts for transaction banks:



- 
- ◆ can we provide predictions of our basic financial numbers, turnover and profit for starters, from just our activity data? If we can do that we have a proper control system for our financial systems that will help us to ask better questions, e.g. if we made more profit on a day than we expected from the activities we undertook, then why?
  - ◆ can we estimate credit risk as well as or better than our credit systems from purely outside information? Why not issue such a challenge to an internal team?
  - ◆ can we model all of our transactions against all of the trade finance we provide? Can we attach our transactions to ship, truck, or aircraft movements? From this, could we work out new relationships with carriers or clients?

These are just some thoughts to get going. The core point is that Big Data requires transaction banking management to think more, not less. Herodotus stated, "Great deeds are usually wrought at great risk". There is the potential for great deeds, but transaction bankers will have to take risks. And that is the essence of science too. Every great theory is great because it can be falsified. Management need to adopt the same attitude and use Big Data scientifically. "Hypothesize, model, test" is not obsolete. Skim the pretty pictures sure, but use Big Data to test Big Theories not fill powerpoint presentations.

[1,710 words]

*Professor Michael Mainelli is Executive Chairman of Z/Yen Group and Principal Advisor to Long Finance. His latest book, **The Price of Fish: A New Approach to Wicked Economics and Better Decisions**, written with Ian Harris, won the 2012 Independent Publisher Book Awards Finance, Investment & Economics Gold Prize.*