Model free data

In most of the world "Big Data" is becoming less credible: in 2013 the Google Analytics flu story [1] was shown to have been exaggerated; Gartner now has it "sliding into the trough" in their "hype cycle"; and the issues of integration have hampered many initiatives started in the heady days of 2012, which are still not showing any benefit. Yet within the upstream Oil Industry, despite the downturn in activity forced by the oil price, there still seems to be optimism about it. Possibly some of that is being driven by vendors and consultants whose revenues depend on "Big Data" projects (and, don't get me wrong, if you are looking for a consultant to advise you I would always be happy to discuss this, or any other, topic). In the past I have made no secret of my scepticism about whether "Big Data" will ever deliver much benefit to oil company technical analysis. There are some domains in which it could potentially be a good approach, if only the right data could be assembled. Unfortunately current oil company practices make collecting suitable data so expensive that at least for now it is impractical, and I don't see any chance that this will change for the immediate future.

A few years ago I read a "Wired" piece that predicted that the collection of large volumes of data would negate the need for human expertise. In other words that computer algorithms analysing heaps of raw measurements would become better at identifying the trends and underlying drivers than people. At the time I thought this was distinctly daft, and all I have seen since then has just strengthened this conviction. I suspect (and hope) that this is a rather extreme position is well beyond what most advocates of "Big Data" would claim, but in a few recent pieces about applying the technique to E&P data the same underlying misconception has been apparent.

If it was possible to collect data without following some kind of theory then this level of benefit might be achievable. In reality all data has to be placed into a context to be recorded. Every piece of data must be kept "somewhere" and the act of deciding which particular slot to put it in implicitly defines what you can do with it. This is obvious (surely?), but yet I still see articles on the web that suggest big data obviates the need to think hard about what the goal is, where the original data comes from and how to integrate driller's data with that from the production engineers.

Generating business value from technical data is about asking the right questions, integrating data from different domains and getting people to do the work, there is no such thing as "model free data".

---

[1] See "The Parable of Google Flu: Traps in Big Data Analysis" which is available for download from http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf