Little prospects for big data?

The oil industry is a "risk" business, the lack of certainty about the exact configuration of the subsurface means that even the most well explored area can deliver a surprise. Oil companies do all they can to mitigate this risk, they analyse all the available data, they explore in partnerships and the larger ones play the numbers game to spread rewards across a portfolio of different assets.

In his book "Fooled by Randomness"[1] Nassim Taleb discusses the way that financial market traders are rewarded. He points out that if, as seems reasonable, the difference between a good and bad trader is that a good one is slightly more likely to have a good year, then after a few years the traders that have consistently performed best are overwhelmingly more likely to be the lucky ones, rather than the good ones. Suppose we want to identify a single "star" performer (with, let's say a 66% chance of having a good year) amongst 15 "average" traders, when we look at 4 years trading more than half the time the good trader will have faltered and one (or more) of the average traders will have clearly better results. Of course the individuals involved will claim that their success is based on innovative analysis, an expert view and better than average intuition.

In the long term all the probabilities even out and the best performers become apparent, but in the long term we're all dead. In other situations the dependence on luck is less pronounced, for example the fact that tuna have to be fast in order to catch pilchards, this is much more toughly tested, each tuna gets thousands of attempts to catch their prey every year, any slight advantage that an individual fish gains (by being a bit faster, a little more agile or a tiny bit better able to anticipate) translates into a significantly increased chance of survival.

"Big Data" has received a lot of publicity over the last few years, I'm not sure I really understand what that phrase means, but my perception is that it is focused on gathering vast amounts of raw data and using statistical techniques to identify previously undetected patterns. This obviously works well with things like shopping habits where there are millions of transactions to analyse and the underlying shape can nudge through the probabilistic fog. Most oil company data and decisions are not like that, even where the data that is "physically big", such as seismic, it has a regularity that ensures an experienced geophysicist will outperform any algorithm every time. A typical oil company makes relatively few "major" decisions, such as selecting which prospects to develop, which wells to drill and where to commission a new seismic survey. There are few opportunities for Big Data to shine.

[1] ISBN-978-0141031484 personally I found this earlier book much better than his later (and more famous) "The Black Swan: The Impact of the Highly Improbable"