

# The Flood of Data

Every so often there is an article, presentation or whitepaper about the fact that oil companies are being drowned in an every growing volume of data and that if nothing is done they will soon be unable to cope with this deluge. But is that really what is happening? I suspect that this is a case where cause and effect are exactly the wrong way round. If data cost nothing to obtain, store and manipulate how much of it would oil companies have? In the competition to understand the subsurface every fact, measurement and inference has the potential to nudge our comprehension ahead of our rival's. Surely, if data was completely free everyone would have an infinite amount of it.

In reality, of course, nothing is free. Every piece of data has a cost to obtain it, a cost to store it and a cost to find it. Many oil companies have a formal process, called something like "Value of Information" that guesses what impact new data would have and justifies the significant cost of, for example, shooting a new seismic survey, by the eventual financial benefit it will deliver. Those calculations focus on the "big ticket" items, expensive activities carried out by external groups to deliver identifiable chunks of data, but most data is generated internally and is never seen by the "Information Management" team.

In a perfect world every oil company would assess the total costs that each set of data imposes and the corporate benefit it delivers in order to decide which data is worth investing in. In reality even benchmarking the activities of the "Information Management" team is a challenge, and the majority of "data management" activities are carried out by users (although they wouldn't call it that of course).

One consequence of Moore's law is that, within a fixed budget, the volume of storage and number of available compute cycles doubles every 18 months. So you should expect data volumes to increase exponentially even if you do nothing to improve, as can your competitors. The executives are consistently striving to make decisions as efficiently as possible, if they are not driving the "data handling" part of the business up to its limits their shareholders would be justified in asking some awkward questions. In most oil companies this year's "data handling" budget is determined by taking last year's budget and asking for savings, the fact that executives don't comprehend how this imposes limits on the total "information throughput" rate of the organisation is hardly their fault. It has to be up to us, as data handling experts, to explain that to them.

My guess would be that the volume of data in any company is mostly determined by the activities users and data managers employ to control it, in other words the volume of data an oil company holds will expand to fill their capacity to handle it (and then a little bit more). So when the volume of data you're handling doubles celebrate this as a sign of your past success rather than a portent of your future doom.

