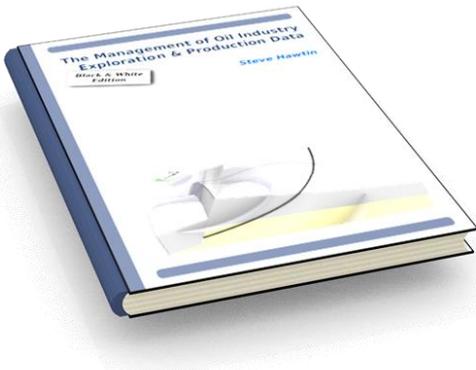# Body of Knowledge

What are the essential things that we should use to accept (or reject) a candidate to be a data manager in an oil company? Of course my personal view was covered in my book and although it was written a couple of years ago I would still claim that any competent manager of technical data should be familiar with all the things I described there. But, of course, however much I would like every data manager to have read it even I have to accept that is unrealistic. So what would I include in my list? There are a number of topics that I feel should obviously be included: the data categories oil companies use; the value of data; DAMA; data quality; maturity metrics; the specialist technical terms; and how data management fits in the larger picture.

The range of technical data employed in oil companies, seismic, well headers and paths, drilling data, petrophysical data, reservoir data, spatial data, facilities engineering data and document management, each has its own profile. A good data manager should be roughly familiar with all the main data categories, how they are employed and some of the key challenges in handling them. I would also expect that any experienced data manager would be an expert in at least one of these data groupings.

The value that data and its handling delivers to the company and some of the ways this can be estimated is, for me, also an essential component. It's not very long since I went into some detail on this topic in these articles, so I won't reiterate (look at old articles if you want to see what I think). The 10 DAMA functions and how they each apply to E&P data is another topic that I've beaten to death (although the anticipated update to the DMBoK will change to 11 functions, so I might have to revisit soon). The various dimensions of data quality (such as completeness, uniqueness, consistency, timeliness, measurement accuracy and reasonableness) can each be used to determine and perform mandatory or indicative quality checks. Of course any set of dimensions you select has potential overlaps and needs to be clearly documented (and also cover the whole range of checks that domain experts are likely to articulate). Enough has been written about maturity metrics (by myself and many others) that I would hope everyone is now familiar with the concept. If you feel you need more you should track down some of Jess Kozman's material (like his YouTube interview).

I suspect there would be little debate among any of the "old hands" about the importance of any of that material. The specialist terms, however, would cause a bit more discussion. Terms such as "corporate repository", "working area", "data approver", "definition owner" and the distinction between "publishing" and "loading" have a range of nuances (and just don't get me started on "metadata" or "data owner"). Suffice it to say that writing down a definition of what these terms mean is always a good idea (and never ever assume anyone else uses them the same way you do). Finally (and perhaps just as controversially) I would recommend having some kind of view on "Enterprise Architecture": in particular how information works, as distinct from business processes on the one hand and applications/ infrastructure on the other. But, and I don't know if I mentioned this already, the best way to review what every data manager should know is to read my book.