

10 things I hate about subsurface data management

- Published on August 12, 2019



Jane McConnell

Practice Partner Oil and Gas at Teradata

7 articles Follow

Talk about big data analytics in the subsurface domain and what's the most common response? "Can't change." "Won't work." Sure, there are some perfectly valid historical reasons why we in the subsurface data management community look after our data the way we do, but isn't it time to break from tradition? Why? Because new and evolving E&P business processes are driving whole new workflows and data architectures across much wider value chains, with shorter time frames and stronger cost control. If we don't change our practices, if we do not build big data analytics into our data management strategies, we won't be able to deliver what the business needs.

Here then are the top ten things that I (and lots of others) hate about subsurface data management today, things I see as barriers to change.

1. Data silos

Seismic SEG-Y files go in one database. Well logs in DLIS, LIS, LAS in another. One for raw log files, and another one for corporate spliced and composited logs. Core data, yet another database. Same with core photos. Geochemistry – separate. Biostratigraphy – yup, you guessed it – different again. Why? Did we miss the lesson on modelling multiple types of data into a single integrated data model? Did we miss the lesson on high performance databases, where it's OK to store millions of rows of data in a single database table?

Keeping data in separate systems with separate indexes, separate master data management issues and often separate physical hardware, only means extra work, master data management problems, and unnecessary hassle when we try to bring the data together so we can analyse it as a whole.

2. Application silos

What could be worse than data silos? It's got to be: storing data inside proprietary application databases or data structures. What's the result? Pools of data that can only be accessed through an application API. Too many reasons why this is a bad idea, here's just a few:

- The data storage strategy is based on requirements for application data access, not on data management principles. Governance and lineage are not a priority
- Most applications still store data sets as files or blobs for performance reasons, limiting your ability to use this data for analytics
- You are locked in to which applications you can run against the data store – killing your ability to choose “best of breed”
- Requiring access through an API prevents you from using mass-market SQL-based tools (visualisation, data quality, MDM) to manage and access the data
- You are beholden to the application vendor and the changes they choose to make to the data model across versions – and you need to implement their upgrades, which are often costly service engagements

3. Library style data management

OK – I understand the history. Our main role used to be to catalogue tapes. But now that the data is often kept on-line on spinning disk, why are we still cataloguing files as closed entities, “black boxes”, rather than cracking the files open and loading the contents into a data structure where we can work directly with the data? Is it because we always did it this way, or because we really don't believe the alternative is possible? [See 1 – did we miss the lesson on high performance databases, where it's OK to store millions of rows of data in a single database table?

4. Project, corporate, or master?

As if we don't have enough silos with our project databases, we thought we should add some more. Today's E&P application vendors hawk a suite of solutions – the problem? Each only deals with *a part* of the data management problem. Sometimes the split between products is by design, but often it's an accident - the result of acquisitions or solutions developed for a single company. E&P application vendors are **not** experts in data management. Don't buy the marketing story about the reasons why you need a separate database for this one data type – they are just excuses.

5. Never fixing the data

When we find incomplete or incorrect headers, when we track down that the wrong CRS conversion has been used, or we finally identify the source CRS – why not fix the data once and for all? Why are we happy to preserve the mistake for posterity in our archived dataset? It's one thing to have provenance and lineage, a system of record – but it's another to refuse to fix the metadata or master data because “it's the original”. Horizontal data management solutions have many options for maintaining a system of record.

6. Big data vs “lots of data”

“In the Oil Industry we have always had big data.”

No. In the Oil Industry we have lots of data. Normally in a cupboard, sometimes still on tape [See 3]. Some of it loaded (multiple times) into proprietary application silos that control what we can and can't do with the data [See 2].

Big data is still our Achilles heel. Awkward and unwieldy data formats trip us up when we try to run analytics on that data, preventing us from realising the true value of all that costly-to-acquire data in its full business context. New tools and techniques could allow us to do things differently.

7. Decisions by PowerPoint

Billion dollar decisions are made on information presented on PowerPoint that contains no lineage information back to the original data on which the interpretations or models were made. Our use of siloed applications and manual data management makes it very difficult to forensically dissect previous decisions, and learn from our success or failure.

8. It's what everyone else does

Well, **everyone** used to think the world was flat, and that the sun orbited the earth. Why are we so quick to dismiss alternatives for data management that have thrived for decades in other industries? If the O&G company that you benchmark against hasn't adopted it, that doesn't mean *you* shouldn't. History doesn't remember everyone, but it does remember Christopher Columbus and Galileo Galilei.

9. Unit conversion issues

With the amount of scientific data we have, there are literally thousands of unit conversions required. And we need it to be accurate. Which means knowing what units our data was recorded in. Metres, feet? Or for geospatial data – we have eastings and northings and we know the data is projected in UTM zone 32 but was the datum ED50 or WGS84? Get this one wrong and your position could be wrong by 200m - and that is not OK for a drilling target.

We know how important this is – and yet we are content to rely on the conversions built-in to applications, many of which are out of date or incomplete. When you couple this one with 5 [Never fixing the data] and 2 [Application silos], you get this ridiculous world where some people know that you shouldn't use the “convert on unload” to export data from PetroBank MDS if it is ED50 north of 62 degrees and loaded before 2005. And the others? Well, they just have to prepare to fail.

10. The low bar of “not losing stuff”

We talk all the time about professionalising E&P Data Management, but at the same time we consider our role to be somewhere between geodata loading monkeys and librarians, with the low bar of not losing the data. Of course, there is the mundane, commoditised data custodianship

that still demands fantastic domain expertise. But why are we selling ourselves short? We can provide a whole new set of capabilities, routinely used by other industries to add value to the business.

Does this list strike a chord? I'd like to know what you think.

Originally posted on the Teradata International Blog site.