

10 Things We Hate About Subsurface Data Management

JANE MCCONNELL, SENIOR INDUSTRY CONSULTANT FOR OIL AND GAS, TERADATA



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 and forgo building big data analytics into our data management strategies, we won't be able to deliver what the business needs.

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?

Here then are the top ten most hated parts of subsurface data management today that are 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. There are so many reasons why this is a bad idea, here are 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

Okay, we 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

1
Data Silos

2
Application Silos

3
Library style data management

4
Project, corporate or master?

5
Never fixing the master

6
Big data vs. "lots of data"

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?

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 application only deals with 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, track down that the wrong CRS conversion has been used, or 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 off by 200m, and that is not acceptable 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. Couple this one with 5 [Never fixing the data] and 2 [Application silos] and you find yourself in 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.

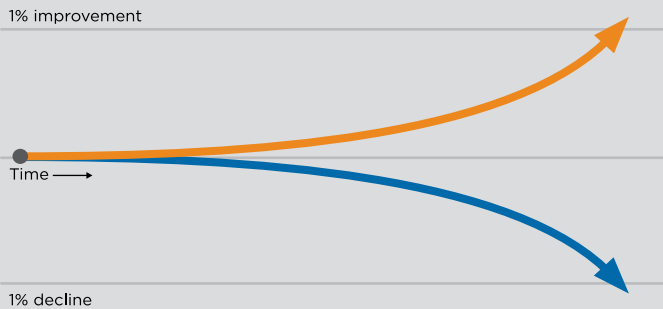


So what now?

Does this list strike a chord? It's a bit scary to admit that you need to make changes. So, let's take it a step at a time. Everyone's heard of the 80/20 rule and Stephen Covey's "big rocks" approach—but have you heard of "aggregation of marginal gains"?

This approach [Figure 1] hit the headlines with the success of Team Sky (Great Britain's professional cycling team) in the Tour de France. When Dave Brailsford took the role of General Manager and Performance Director for Team Sky, he implemented a concept that he called the "aggregation of marginal gains"—making a 1 percent improvement in every aspect of his team's work. He believed these small gains would together add up to remarkable improvement.

Aggregation of Marginal Gains¹



In the beginning, there is basically no difference between making a choice that is 1% better or 1% worse—in other words, it won't impact you very much today. But as time goes on, these small improvements or declines compound and you suddenly find a very big gap between people who make slightly better decisions on a daily basis and those who don't.

1. James Clear, "This Coach Improved Every Tiny Thing by 1 Percent and Here's What Happened".

Brailsford started by optimising the things you might expect: nutrition, weekly training program, the ergonomics of the bike seat, and the weight of the tires. Then they searched for 1 percent improvements in areas that were overlooked by almost everyone else: the best pillow to optimise sleep, the most effective massage gel, the best way to wash their hands to avoid infection – they searched for optimisations everywhere.

Brailsford believed that if they could successfully execute this strategy, then Team Sky would be in a position to win the Tour de France in five years' time. He got that wrong—it only took three years to win the title.

When people talk about optimising company performance, they usually concentrate on the core activities and deliver "good enough" on the rest. But what Dave Brailsford did by optimising all areas for excellence—wherever he found them—he proved that the smallest improvements can have a huge effect. Within data management, this same philosophy can be adopted to have a large effect on the overall performance of the company.

So what's that got to do with subsurface data management? Well, when people talk about optimising company performance, they usually concentrate on the core activities and deliver "good enough" on the rest. But what Dave Brailsford did by optimising all areas for excellence—wherever he found them—he proved that the smallest improvements can have a huge effect. Within data management, this same philosophy can be adopted to have a large effect on the overall performance of the company.

This approach could have unexpected advantages in our cost-focused world. As Brailsford puts it: "You'll get more from a £900,000 rider with a coach than you would from a £1m rider without one." It's the same with data management. Data underpins any business—use it to your advantage.

Here are some one percent improvements you might consider making in subsurface data management:

- **Store your bulk data**—seismic surveys, raw logs, core photos—in a Hadoop system so you have access to parallel compute power right where your data is—no need for separate application servers and the resulting I/O issues.
- **Keep metadata with your data at all times.** One approach we have seen for raw data is storing the raw files together with their operational metadata in key-value pairs in Hadoop so they never get separated. When data is transformed from one format to another, consider tools like Loom to track the transformations.
- **Build a corporate store on an MPP database platform,** so that as it grows you keep linear performance and your beautifully curated data store does not turn into an unusable data dinosaur.
- **Combine data stores wherever possible,** removing the possibility of error through master data issues and alternative conversion engines, while also making analytical queries easier and faster.

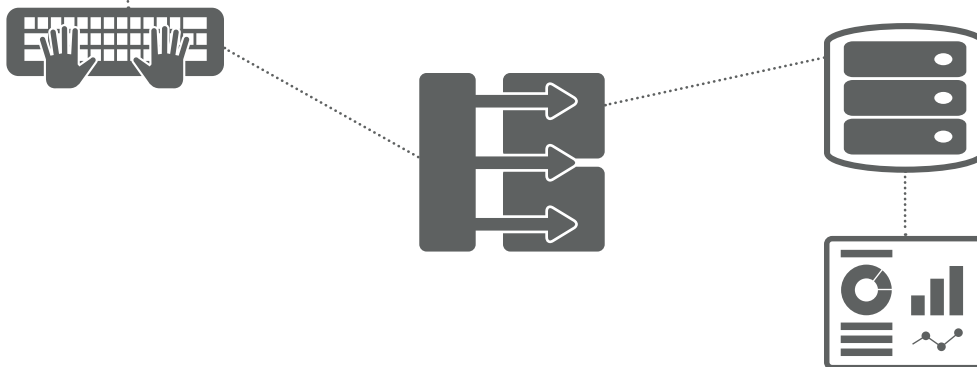
- **Implement a discovery platform**, so you are ready to run any analytics against any data. Then, when the business comes up with an interesting data question that has been keeping them awake at night, you can help them answer it in days, not weeks (or months).
- **Incorporate text analytics** on drillers' logs, operations logs, and maintenance logs, and analyse in combination with the structured data. There are nuggets of gold in these text fields.
- **Automate a constant quality check** on unit conversions and coordinate conversions across disparate systems.
- **Choose an area where lineage is lost today** (user exports data to Excel to make calculations, then emails the Excel file, for example) and implement a better solution.

Like the thinking? Teradata helps companies across many industries, including those in Oil and Gas, get more value out of their data than anyone else. Don't take our word for it—Teradata is positioned as the leader in this space by analysts like Gartner and Forrester, year after year. Talk to Teradata about making those one percent improvements in subsurface data management today. Teradata.com/oilandgas

About the Author

Jane McConnell - Senior Industry Consultant for Oil and Gas, Teradata.

Jane joined Teradata in 2012 as a Senior Industry Consultant for Oil and Gas, based in Norway. She holds a BEng Honours in Information Systems Engineering from Herriot-Watt University, Edinburgh. She has chaired the Education Committee for the European oil industry data management group ECIM, and regularly presents at Oil Industry data management events.



10000 Innovation Drive, Dayton, OH 45342 Teradata.com/oilandgas

Teradata and the Teradata logo are registered trademarks of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions, and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or Teradata.com for more information.

Copyright © 2015 by Teradata Corporation All Rights Reserved. Produced in U.S.A.

6.15 EB8875



TERADATA.