Finding Petroleum

Why this isn't going as fast as expected

Bain - organisational and capability aspects

Using automated data clean-up techniques

Why oil and gas struggles at digitalisation

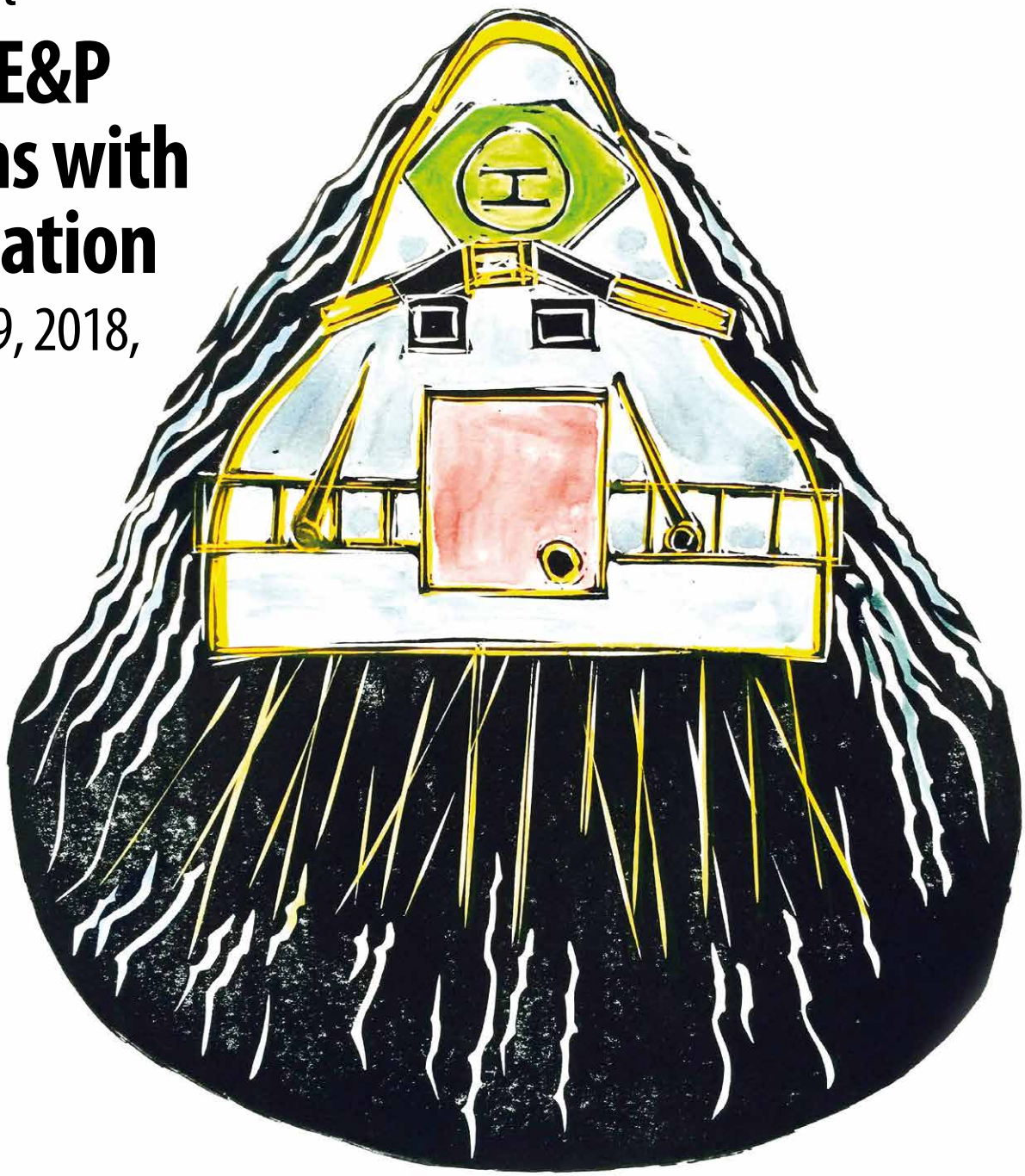
Implementing analytics as a system

Doing more with a data lake

Special report

Solving E&P problems with digitalisation

November 19, 2018, London



Finding Petroleum

This is a report from the the Solving E&P problems with digitalisation event - November 19, 2018, London

Report written by Karl Jeffery,
editor of Digital Energy Journal
jeffery@d-e-j.com
Tel 44 208 150 5292

Finding Petroleum
www.findingpetroleum.com
Future Energy Publishing,
39-41 North Road,
London, N7 9DP, UK
www.fuenp.com

Sales manager
Richard McIntyre
rmcintyre@d-e-j.com
Tel 44 208 150 5291

Conference producer - David Bamford
Report author - Karl Jeffery
Layout - Laura Jones, Very Vermilion Ltd
Photos - Avinga Pallangyo

Note:
Some of the slides and videos from this event can be downloaded free from the Finding Petroleum website event page
http://www.findingpetroleum.com/event/96ebb.aspx

Making oil and gas digitalisation work

Just about all of the digital projects in oil and gas companies today are underperforming, we heard at our November 2018 London Finding Petroleum forum "solving E&P problems with digitalisation." Some reasons are that expectations are wrong, people are introducing technology for the wrong reasons, or we are going the wrong way about it.

"Many senior management in oil and gas companies are asking themselves, 'why is this [digital] thing not working as quickly as perhaps we thought it would,' with no substantial impacts on the company bottom line," said Peter Parry, partner and leader, oil and gas with consultancy Bain, speaking at the Finding Petroleum forum in London on November 19, "Solving E&P problems with digitalisation."

Just about all of the digital projects in oil and gas companies today are underperforming, he said.

A simple list of digitalisation projects an oil and gas company might be working on today might include autonomous robots and vehicles, additive manufacturing (3D printing), internet of things / wearables / sensors, digital engineering and training, virtual and augmented reality, cloud and security, big data / advanced analytics, artificial intelligence, mobile and digital engagement, and robotic process automation.

The three areas furthest ahead in showing results and building new capabilities are probably autonomous robots and vehicles, digital engineering and training, and big data /advanced analytics, he said.

Our forum explored some techniques to make digitalisation work harder, including agile working methods, open platforms, pursuing technologies the entire organisation can benefit from, hiring more data analytics people, using automated data clean-up techniques and language translation engines.

Also not trying to be like Uber, purchasing for practical reasons rather than hype, giving data scientists more access to domain experts, moving more data onto a public cloud (including software applications), implementing analytics as a company wide system, and having realistic expectations and good governance processes for your 'data lake' projects.

digital energy journal

Note: many of the videos and slides for this event can be downloaded free of charge on the event web page **http://www.findingpetroleum.com/event/7bb4f.aspx**

Bain – organisational and capability aspect the hard part

While technical aspects of digitalisation are "relatively straightforward", the organisational and capability part can be "something of a minefield," said Peter Parry, partner and leader, oil and gas with consultancy Bain

While technical aspects of digitalisation are "relatively straightforward", the organisational and capability part can be "something of a minefield," said Peter Parry, partner and leader, oil and gas with consultancy Bain. But "the organisational components and the capability components of this are really fundamental to getting it kick started.

"It's not a one-time decision, it's a generational thing. We change the organisation slowly," he said. "We build competence and capability over decades sometimes. But here is a subject where we need to build those capabilities pretty quickly."

"The organisational components are going to be very fluid. We're not going to end up with hard and fast structures, we're going end up with very dynamic things. But we need to start to move forward without a hard and fast destination in mind."

The big question is what is the right capability and organisational structure to deliver this. "The answer is rather difficult in practise, rather easy in theory," he said.

It is commonly said that people overestimate the impact of digital in the short term (how quickly they will see good results). But they also underestimate the impact of digital over the long term, the sizes of the changes which will be possible.

Digital can be seen as a number of different 'waves hitting our organisation, first with the technology, and then in how it affects people and physical assets.

Benefits

To see an example of what happens when you get digitalisation right, it might be helpful to look at the wind energy sector, where companies have improved the physical performance of their asset based on digitalisation of all aspects of operations, he said.

"Running a wind farm with outstanding digital capability to support and optimise it can add 30 per cent productivity to that asset," he said. "It can improve return on capital of a renewable project from around 8-9 to 12-15 per cent. Some oil and gas companies are taking upstream people into their renewables business and say come and have a look at this," he said.

Peter Parry, partner and leader, oil and gas with consultancy Bain

Mr Parry shared similar ideas at a Rosneft technology conference in Moscow in November 2018, with an audience including an asset manager from each of their assets, and a representative from each of their service and supply companies, adding up to about 2000 people.

For an organisation the size of Rosneft to realise the potential of digital technology, getting the organisation to work together is key, he said.

Agile

One new way of working is "agile," which can be described as a way of working which means "putting things that are not working aside very quickly and focussing your time and attention on things that are working," he said.

The Agile working method was developed in domains such as product creation, but can prove to be "fantastically differentiating" in oil and gas exploration. "You can squeeze projects that took 4-5 years into several months by working in that way," he said.

Agile is "an entire way of working," with people who can focus on ideas, get ideas to the business, and deliver as well.
There can be big resistance to agile ways of working, or people thinking it is nothing new. "It's not how you will work, but about how teams will work, how quickly they will get negative outcomes off the table and positive outcomes moving forward," he said.

It can be difficult managing procurement in an 'agile' world, because you don't yet know exactly what you want. You need to structure the purchase as an 'outcomes' not physical assets. But companies are used to paying for specific machinery, or people-hours. An agile contract might say the outcome needs to be in a certain range, and the payment will vary depending on what the outcome is. A contractor which proves unable to take the project forward needs to be quickly dropped.

People who work in technology organisations can often seem to do projects "for their own entertainment," he said. "This process has to be pretty brutal in sorting the good from the bad. "

Open platforms

Another good way of working is "open platforms". This can mean a reverse of how the industry got competitive advantage over the past few decades, by having proprietary technology or insights, so it can close doors and have "things that we can do that you can't."

"Getting the best out of digital is about opening those doors, being the easiest to work with. About allowing others to build on your platform," he said. Similar to how many software companies are making apps for mobile phone platforms.

"I just finished working with a very large company in their technology division. Getting that group of engineers to think open platform as opposed to proprietary technology is a massive change," he said.

Where do you start?

One challenge is understanding where to start, with so many different opportunities.

A simple list of projects oil and gas companies might want to be working on might include autonomous robots and vehicles, additive manufacturing (3D printing), internet of things / wearables / sensors, digital engineering and training, virtual and augmented reality, cloud and security, big data / advanced analytics, artificial intelligence, mobile and digital engagement, and robotic process automation.

"This is a typical portfolio of digital projects in E&P company. They are technically fascinating and can consume an enormous amount of time and resource," he said.

But in terms of business results coming out of the back end, just about all are underperforming.

Of these 12 sectors above, the ones with the best results are autonomous robots and vehicles, digital engineering and training, and big data /advanced analytics. "These three areas you would expect to be furthest ahead in having

deployed digital technology, in having seen results, in terms of building new capabilities and sorting out organisations wrinkles, bottlenecks."

Autonomous vehicles and robots are increasingly used onshore and increasingly offshore.

Digital engineering and training is proving to have fascinating potential. One idea for training could be for offshore workers flying by helicopter to start their shift having training via their headsets on the flight, getting an update on what has changed while they were away. "Training doesn't mean just sitting in a room and listening," he said.

Big data, advanced analytics and AI prove "easy to say, pretty hard to deploy at scale," he said.

Ease of implementation of these technologies is usually a matter of capability, if you have a workforce able to use that technique, information, or changed way of working very quickly.

Meanwhile oil and gas companies are clear about what impacts they want to see, such as a big change in their HSE performance (which can be simply from taking people out of hazardous environments), and reduced capital costs of 20 per cent.

One-off results

When oil companies cite specific results from digital technology, he said, they typically refer to one-off examples, rather than across their whole company. So they have reduced costs by 60 per cent but on just one project, or seen a 30 per cent production increase but on just one field.

This is equivalent to a bank saying that its customer satisfaction has improved by 60 per cent, but then saying "the customer who is more satisfied is that customer over there," he said.

This focus on performance over just one project means that the organisational elements and capabilities get left behind.

The entire organisation

The only way to get a big impact from digital is if the entire organisation does it. "I'm rarely going to get a big impact from a few folks doing something," he said. One of the biggest struggles big oil companies have is with "speed and scale" – they have some results, but it isn't being rolled out fast enough or having a big enough impact.

"It is only when you are investing heavily behind capabilities you're going to get a good return on that investment, and maybe a significant competitive advantage. "If you don't have these things right you're wasting your money, it's a hobby, and it's never going to have a substantial impact."

Three organisational elements you need to make things work at scale are some kind of mission or strategy, some targets and governance, and operating models which include digital.

The strategy is needed to get the necessary resources committed to the digital initiatives.

The governance means there is clarity over who is in charge of the project and responsible for it. Sometimes this is shared between different parts of the organisation, such as the upstream business, regional technology centres or an asset manager.

The operating system needs to be updated to include digital. Many companies have operating manuals but have not updated them. "You cannot find, in a major oil and gas company management system, a description of how to manage drone operations," he said. "It is somewhere under aviation."

The capability isn't necessarily all built internally, you can buy it in, for example by hiring data scientists.

Many people believe that the relevant business departments of oil companies, staffed full of engineers, must have the competence and capability to handle digital technology.

But while traditional oil and gas technology might change every 6 years, this sort of technology changes every 6 months. "So if you're not focussing a lot of attention on how the outside world is adapting you are basically implementing yesterday's solution."

How to deploy

It is often not the choice of technology which is most difficult, but the way to deploy it – and whether it should be pushed by the company centrally or by individual business units.

If technology does not have a particularly big impact, it might be better just to let the relevant business department decide whether or not to implement it.

If it only involves one business area, then again it should probably be implemented by that business unit.

If it does not have any disruptive or negative implication on anyone, then again the relevant business unit should be involved. And if the business unit has the capability to run the project then it probably should.

The company's headquarters should get involved in other cases – for technology with a very large potential impact, which may involve multiple business areas, some negative impact or disruption to some people, and new capabilities.

Out of oil and gas 33 digital projects Bain considered, it turned out that only 7 of them should have been handled at a corporate / headquarters level.

More data analytics people?

There is often a big shortage of data analytics people, if it is defined as people who use business information to try to improve the overall performance of the business. About 0.1 per cent of oil and gas employees are doing this. Engineers, geologists and geophysicists are all trying to get value from data, but with a different objective.

By comparison, about 1 per cent of employees in the automotive sector are working with data trying to improve business improvement, and about 7 per cent of employees of tech companies such as Amazon, Google and Netflix, and this number may be too low.

"I would argue, 0.1 per cent is not going to get you anywhere," he said. "If you don't change the way of working to realign with the way digital is going to provide capability, you won't get anything out of it, it's as simple as that."

You also need the IT support functions able to work with the new software, and not always trying to catch up. One example is a company which bought a big piece of software, but their IT system could not enable it to be used in more than one location at once.

One oil and gas company, a "mid-sized European player", was looking at the best way to set up an analytics team. Bain showed them what it would look like, taking best practise from other industries, and suggested there would be a customer or business function, a business analyst, data scientist, chief analytics / data officer, a data architect, data engineer, data analyst, data visualisation engineer, IT function support, service providers (e.g. cloud), platform service provider, digital ecosystem partners.

The analytics person is in this team, but sitting around a large infrastructure of people who can develop the system and take it forward. "You don't have to hire all these people but you have to have access to them, have them plugged into your system," he said.

Finding Petroleum

Using automated data clean-up techniques

With data being generated so quickly, organising data manually isn't feasible any more, you need a machine to help. Waclaw Jakubowicz, managing director of Hampton Data Services, shared some tips

In the past, or up to the present day, it was possible to manage or clean data manually, as with physical libraries.

But now data is being generated so fast it is impossible to do it manually. So you need a machine assisted process, explained Waclaw Jakubowicz, managing director of Hampton Data Services.

For example, machine learning tools can analyse documents to see which words occur most option, and try to classify it automatically. Another technique is to link data to objects, and then classifying the objects.

They can see which data appears to be related to other data, from looking at references in the headers / metadata.

You can get a sense of the general patterns of data about production, engineering, economics, and field development.

Once you have a sense of how data is created, you can see which data is missing, and then try to find it.

Machines can analyse data much more widely than people can – people typically just clean up the data they need to work with, Mr Jakubowicz said.

A challenge with any data clean-up project is that new data is being created all the time, which needs to be stored so the system understands which wells, assets or subject matter it relates to. Managing new data also requires active data management work. "You cannot rely on users to nicely file a file. They'll make 20 different versions," he said.

Managing PowerPoint files as also part of today's data management work, since they are typically made at the end of a project to summarise everything, with investments made as a result of them.

Case study - Reach Energy

Hampton Data had a data clean-up project with Reach Energy Bhd of Kuala Lumpur, which had bought a controlling interest in Emir Oil LLC in Kazakhstan. It came with a great deal of legacy data.

The database was multilingual, including with material from Beijing and research institutes in Kazakhstan, all poured together. Some data was in Chinese, some in Cyrillic character set. The main dominant language and character set was English.

Waclaw Jakubowicz, managing director of Hampton Data Services

The data had many co-ordinate problems, and poor notation about what comes after what. A number of different data management companies had tried to improve it.

A first step for Hampton was to move the data to its own server in the UK. A separate copy of the data was kept in Aktau, Kazakhstan, synchronised with the data store in London. This means there is a complete backup in both locations. This covered both new data and legacy data.

Then it started a number of processes to rationalise and clean up the data.

An initial problem was understanding well and place names. Some wells were given multiple names (or aliases), or their names are spelt in different ways in Cyrillic. There can be files named in English, Russian and Chinese in the same folder. "You have to be multilingual to get your head around that, he said.

Hampton Data has developed its own translation tools through its work in different countries over the years, so it can auto translate file names from Russian and Chinese into Latin characters. The headers can also be auto translated – but with the formatting maintained.

Often the file name will itself indicate what the file is about, for example "core data from xyz well", or "PVT analysis". This means that English speaking engineers trawling through the data find it laid out for them nicely.

Hampton Data works with a company called XTM, which specialises in managing technical documentation, and also works with many large automotive companies. It gathers libraries and vocabularies specific to the industry, something Google Translate does not do. Documents can also be translated for other users, not necessarily into English.

Nephin Energy

Another client is Nephin Energy, a start-up company based in Dublin, Ireland, which acquired a large gas field offshore Ireland, formerly operated by Shell UK. The investment funds came from a Canadian pension fund. Nephin produces 60 per cent of Ireland's natural gas.

The data was very organised, as you might expect from Shell. But the volumes were very large.

It would have taken a few months to do a data audit manually. Hampton was able to do it in a week with automated tools.

Nephin runs with a very low number of employees, and is outsourcing as much work as possible to outside consultants. It uses Microsoft Azure for its IT infrastructure, and would like to have all of its data and applications on there.

One disadvantage of Azure is that "every time you look at data, move it about, you get an invoice hitting you," he said. "It is an unpredictable beast, no-one knows what it will cost them at the end of the day."

The company has moved data to the cloud in the same format as it was when they acquired the asset, they are not re-arranging any folders. Hampton provides a virtual "data custodian" system which runs semi autonomously, keeping the data organised.

It would be helpful if the applications and data could be stored on the same cloud infrastructure. But big subsurface software providers typically only want their software to run on their own cloud, which makes it tricky.

"If you want to bring your own bit of software like Hampton Russell or something else, it is not exactly encouraged," he said.

There can be some flexibility, but it generally ends up that the larger the oil company, the more leverage they have to dictate which cloud will be used.

There are many smaller software companies who would like to run tools together with other software, including subsurface time depth conversion software, various simulators, petrophysics applications. But they can't, if they don't have access to the same cloud that the bigger software is running on, he said.

For example, one start-up company called Anteaus Technologies is looking at applications for wells, such as log interpretation and geomechanics. They have developed applications to work on the cloud.

Finding Petroleum

Teradata - why oil and gas struggles at digitalisation

If digitalisation is using data to drive your decision making, then perhaps that explains where companies are going about it the wrong way, if they start by trying to be like Uber, said Duncan Irving of Teradata

Duncan Irving, oil and gas practise partner with Teradata, defines digitalisation as when you can use data from your various operations and processes to drive your organisational decision making. With the right sort of data, it becomes a competitive capability.

So far the oil and gas industry is nowhere near as good at this as many other industries. In order to work out how to improve, we could try to understand why that is.

The industries which have done well in digitisation gather large amounts of structured data as an "exhaust" from their normal processes, which they can analyse.

The oil and gas industry by comparison has data generated from decades of use of physical equipment and assets, with very little structure of standardisation across different assets, and data often stored in silos, not integrated together.

The oil and gas industry has also not been under so much pressure to improve its competitive capability. Even in the unconventionals sector, which has high cost control, the biggest success factor could have been how quickly you can pay off your loans," he said.

When the oil price was high, the industry was not under much pressure to improve performance, and also the lack of proven use cases discouraged others. There were also concerns about negative impact of analysing data, such as security impacts, or a risk of damaging critical data.

Now, the industry is under pressure to improve performance, but many oil companies get in the wrong direction by saying they want to be like Uber. But Uber's power comes from changing the way a product is sold, Dr Irving said. The oil and gas industry is not planning to do this.

But analytics can still provide plenty of value. For example, if we can standardise data for all the well plans we ever drilled, then we can analyse data and maybe relate it with data about the well's performance over its lifecycle, and understand why some well plans lead to better wells. The well planners have a better understanding of their process, and which aspects of the well planning process are most critical in terms of lifetime performance.

The oil and gas industry might be better off

Duncan Irving of Teradata

trying to understand the actual decision making and interactions which happen, he said. For example, people might be trying to understand whether one well is similar to another, and analytics might help with that. Or there might be better ways to gather data when multiple companies are involved in a project, for example multiple drilling contractors on one project, who could have a bigger and deeper data communication.

Another problem is how much the industry is motivated by hype in how it chooses technology. Today it is possible for an analytics company to get a meeting with an oil and gas company just by saying they are making 'machine learning', because people have basically been told to get a machine learning project by their management. "People are that shallow," he said.

Many software companies are just putting a 'machine learning' button on their software.

"That's not serving either the organisation or the industry well in understanding how to use these new capabilities properly," he said.

Sometimes when companies say they want machine learning, they are really saying, can we have some cool stuff. What they really need might be "pretty simple stats."

For example, a desirable output could be a simple cross plot graph, showing how one piece of data varies with another one, which reveals the system works differently to how the experts have always believed.

Teradata did a successful project for Siemens with rail locomotives, combining sensor data from trains with operational data, and then being able to make predictions about when various components would fail in future. The project could be considered more data management than data science, he said.

Misunderstanding data science

There's a massive misunderstanding in the oil and gas industry about what data science and data engineering actually mean, he said. There's nothing fundamentally wrong with these disciplines, but you should not expect a data scientist to have the same understanding of drilling that a drilling engineer would. You have to sit a data scientist next to a drilling engineer with 20 years' experience and let them get on with it."
A data scientist is unlikely to be particularly good at preparing data and understanding data quality – and also knowing what a 'system of record' can do.

Data scientists are usually good at statistical analysis of data, which is something most engineers don't understand. But they are less likely to be good at vector calculus, which can be important for physics based simulation.

"We are expecting a lot from people with a PhD in maths – but not the right flavours of maths," he said. "I've seen that go very badly for them. They haven't been re-wired. But it's not their fault, it's our fault."
Sometimes recruitment consultants overpromise when selling candidates to industries, saying they are skilled in machine learning.

Machine learning can be considered part of artificial intelligence, which has been under development since the late 1950s. There is a lot to know about it. Data mining has been under development since the 1980s, looking for relationships and patterns in data which can provide some useful insight. So there is a lot to know about it.
If you have a big data clean-up task, you are probably better off giving it to one of the data management companies which specialise in it, rather than giving it to data scientists.

Locked in software

Much of the technical data in the industry is locked in specialist software applications, provided by service companies.

This could be attributed to many years of oil and gas technical experts finding that they could not get what they needed from their in house IT departments, and so they asked a service company to provide their software and databases as "application suites", and a data management company to manage the data.

The IT departments "didn't understand what a workstation was for, or what those application silos did, or how to maintain them so the users could work with them wherever they were in the world," he said. This move led to data silos, "both physically and culturally".

We have got there because we preferred to "buy" rather than "build". "No-one got fired from buying a few more licenses from your favourite service company," he said.

It means that the only way to integrate software from different software packages is to export it into Excel.

One way out of this is for more data to be stored on the cloud, which means it can be accessed more easily by other systems.

Data clean-up time

It is typical for data clean-up and integration time to take far long than the actual analytics work, Dr Irving said. In one project, the geophysical and geological team took two months to integrate all the data they had about new acreage, then three days to use it to work out where the 'sweet spots' were for drilling.

In other project, to try to rank different options about how to plan and complete a group of wells, it took four months to standardise the data about past well plans, so they could be compared.

Another project team of five data scientists spent six months trying to integrate data from drilling operations and geological/geophysical departments. "And these data scientists are not cheap," he said.

"It makes data scientists feel a little bit sad to be doing something as mundane as this. They come into an industry, full of excitement, and you've got them effectively sweeping the digital floor, it's heart-breaking. Then they go off and work for a bank."

Business, operations and subsurface

The oil industry IT landscape today is split into three "islands" – business, operations and subsurface.

The 'normal business' island, includes business planning, SAP, financial systems, e-mail. The 'topside' or operations island, is where people manage facilities, plant, and safety. Then the subsurface island, is where geologists and petrophysicists work. All of their data is very different.

Analytics is very good with "Normal business" data, and has been working with it for decades.

The operations people do a lot of their own analysis, although they don't have a lot of standardisation.

The subsurface world is full of data with different units, different co-ordinate reference systems. It also does physics based simulation on subsurface models, so analytics within the subsurface world is very hard.

If you have a project which requires data from all islands, it means you have to talk to people in corporate data management, perhaps technology and research, probably exploration departments.
"It is really hard getting a mandate to do even very basic cross domain analytics."

A business process for analytics

Teradata suggests to its clients that a process for doing data analytics should start with a vision which is part of their central corporate strategy. It should include a description of what the company will do, and estimate of how much value the project can provide, so people can see why the project is happening and what they have to do.

Then when analytics people are trying to gain access to various data stores to set up cross functional analytics, the relevant access is more easily provided.
Then you build an architecture to make all of this happen, and develop your more formal analytics capability. "You become digitalised culturally and organisationally," he said.

Companies which just put data scientists on a project asking them to find some value will eventually come into problems, because the understanding data scientists can get from any data is limited, without domain experts also being involved.

Finding Petroleum

Tessella – Putting the 'data' in data analytics

Oil and gas companies are excited about the potential of data analytics. However, they struggle to move from a promising idea to something useable in everyday operations. The problem, says Dr. Warrick Cooke, consultant with data science company Tessella, is the data being fed in to their models.

There is little doubt in oil and gas that data offers huge potential to improve efficiency and safety and save money. There is also little doubt that it is mostly failing to do so.

A major reason for failure, says Dr. Warrick Cooke, consultant with data science company Tessella, is a focus on data tools and models, at the expense of the actual data itself. Garbage in, garbage out, as the old computing adage goes.

There has recently been an explosion in easy-to-use data tools, such as Microsoft Azure, says Dr Cooke. These are extremely user-friendly, and push users to be hands-on and try things out, allowing quite powerful data and machine learning models to be built with relatively little experience.

Users can quickly come a long way with these tools. There are lots of simple tasks that they do well, and they are great for proof of concept models built on well understood test data.

"But they are quite formulaic, and they don't encourage good practice in ensuring results are repeatable when models are applied to messy real-life production data", Dr Cooke says. The result is models which work on test data, but are not fit to be released into the wild.

Dr Cooke makes an analogy to the early days of Visual Basic. "It opened up application development to a much broader audience, but many of these would then break once deployed. Eventually companies learned that making these applications a long-term success needed qualified software engineers."

Get the data right first

Tessella has a history of working with the oil and gas industry to develop models and curate data. "We often find data is the biggest sticking point," says Dr Cooke. "It can be fragmented, incorrectly labelled, missing information such as time or location, or not properly indexed."

He gives an example of an oil company looking at drill readings to analyse drill team performance.

"Often data will not have consistent naming conventions, so two comparable pieces of data are recorded differently," he says. "Equally, different data can be named identically. One company comparing asset performance was using the same Well ID across different regions. Our team needed to update the data before the model would work."

The list goes on. Data is captured in different formats, sometimes even as scanned pieces of paper. Metric and imperial units are mixed. Data is missing; Dr Cooke tells of a project using sensor data, where one sensor was down for half an hour, creating a gap in the time series. Models built using ideal datasets can't deal with these inconsistencies.

If something's worth doing

Models need good data to get good results. This means developing a system for naming things, and agreeing consistent data formats for wells, sensors and equipment. In most cases, it means considerable changes to existing data and data collection methods. This takes time and effort.

Where data is missing, domain experts should assess what it should look like. A data analyst may be able to tell you what they expect it to look like, based on the past patterns, but this is risky. What if the absence of data was caused by an unexpected event? It is often the gaps that represent the most important information for training models to recognise warning signs. Domain experts have the contextual knowledge to fill these in.

Dr. Warrick Cooke, consultant with Tessella (right)

Even with good practice, real-life data is rarely perfect. Good models should be designed to cope with the unexpected. Problems such as inconsistent units or missing data can be overcome, but only if the problem has first been identified and the model trained to deal with it.

Just as critical is testing the model on less than perfect data – of the sort it will encounter in the real world – to see how it performs. This allows problems to be identified and modifications made, either to the model or the data - to ensure it delivers meaningful insights.

Testing should be ongoing. Expanding the model to new assets will bring new data problems which need to be factored in. This is true of any change, including when data sources such as new sensors are added to existing systems, or updates made to the model.

"Building a good model is important, and modelling tools can be a good starting point to test ideas," concludes Dr Cooke. "But if you want a model that works on real-world data and scales across diverse assets, you need to ensure data is properly curated, and models are rigorously designed and tested."

Finding Petroleum

Doing more with a data lake

The idea of a data lake, a central depository for information, is popular in the industry, but many companies don't get the value from them which they expect. Dave Camden, IM consultant with Flare Solutions, shared some experiences

The idea of a data lake, a central depository for information, is popular in the industry, but many companies don't get the value from them which they expect, said Dave Camden, IM consultant with Flare Solutions.

Data lake is a technology which people get excited by, including a new generation of people coming into the industry. But it is also basically just a file store, and getting value from it requires good data management systems.

These are problems which many people have decades of experience working with. "We're still having to think about these things we've always had to think about," he said.

Data lakes can get very large, some as big as 3-5 petabytes (one petabyte = 1000 terabytes or 1m gigabytes). They are usually cloud hosted, and designed for fast data access. The data is usually in its original state (files or data objects). Some people have described them as "a place you put data until you decide what to do with it," he said.

Dave Camden, IM consultant with Flare Solutions

People usually start building them with an empty data store, then they put in folder systems, copy in their files, and design data 'feeds' for new data to go in. "You are buying a bucket to put your stuff in, a file system," he said.

A data lake can be a precursor to analytics, if you need to gather data together first. It is a way of taking data out of data 'silos', making the company data available to everyone in the organisation.

Putting data in the data lake is cheap to do, particularly as the data does not need to be converted. But if there is cleaning, formatting and structuring involved, it can take a lot of time, even if you use automated tools, he said.

It is similar to when people running a physical library ask for a hard copy of every document to keep in the library. It is easy to collect everything, but harder to get value from the documents once you have them. "A data lake is not for everybody," he said.

Some applications may be better off using traditional data structures, such as the data warehouse. "They are still in their infancy, and part of a solution but not an entire solution."

If the data is going to be brought into a software application for a certain task, it will probably need a rigid structure, so good data management is critical.

Wood and trees

When managing data lakes, it can be helpful to recognise that some people want an overview picture, some people want a detailed picture. This is analogous to some people seeing a wood where others want to look at a single tree.

The "wood" approach, perhaps for senior managers, might include business intelligence dashboards and information catalogues, looking at the entire data set.

The 'trees' approach, perhaps for technical specialists, might include doing a search for specific attributes, looking at the complexity of individual objects.

Meta data can help people who work at both levels, giving context around the data for people working at a higher level, and guiding people to the right information at the 'trees' level. The metadata should ideally tell you the source of the data and the processes it has been through.

Exploration, development, production

A first phase of building a data lake could be considered 'exploration', trying to understand what value you have in your data.

This can be followed by a 'development' phase, developing techniques, workflows, thinking about how things work, understanding data flows (including from sensors), doing some data mapping, perhaps a little bit of governance, trying to move towards a "proper production environment", getting a useful business output.

But hardly any oil and gas data lakes make it beyond that to the "production environment" stage, Mr Camden said.

Varied data

All data lakes are different, with different amounts of structured and unstructured information, files in different original formats, different schemas.

There can be more structured information, such as data by time and depth series. There can be more traditional data stores.

Some companies have different data flows, for example you might have a stage before the data lake where you decide whether data might be useful, then clean and structure it, add metadata to it, before feeding into a data warehouse. There are a number of standard techniques for doing this.

Knowledge model

You can make your data lake easier to run analytics on if you have a 'knowledge model' which shows how the various data relates to each other. For example, you already know that different assets have a relationship, or you have a taxonomy structure you use.

You might also want to use machine learning to enhance your knowledge model, if it can work out ways different information is related.

When figuring out how to get a machine to solve a problem with organising data, it can help to first ask yourself "how would I solve the problem." If a person can't solve it, it is "pretty tricky to each a machine how to solve it," he said.

But machine learning requires that the data is in good condition to begin with, which is usually not the case. If people have the wrong context when working with information, they make the wrong decisions, and machines are the same. "There's nothing magical about the process," he said.

Data management

One hope is that the interest in analytics will drive a focus on data quality management and governance, a problem oil companies have had for decades.

Nearly all analytics projects eventually run up against a barrier, that poor data quality stops them going any further, he said.

To implement data management, you need clear strategies, not just developing them but making sure people are aware of them and understand them. "We've seen situations where strategies have been written and the people operating data lakes have no idea what they idea, so the thing turns to chaos," he said. "Management and governance are about making sure things are defined, implemented and monitored."

You might need senior management to support your efforts to improve data governance, and for that to happen, they will want to see that the project is providing benefits to the company.

You also need to think through the different security and access requirements for the different people who will use the data lake. You need standards for metadata and standards for the process for loading, stewardship and delivery of data.

The data lake may contain a copy of data stored in other places, in which case you need a process for managing the duplicates.

Finding Petroleum

What did you enjoy most about the event?

Open discussion and knowledge sharing.
Troika International

Stimulating and informative presentations.
Simon Berkeley, Berkeley Associates

Interesting overviews of the interaction between the Oil and IT industry approaches.
Richard Walker, Cornhill Economics

Because of my interest the presentation by Peter Parry was most enjoyable, very clear, precise and clearly he knows the subject very well.

Presentation from Bain and Company.

The presentation by Bain's Peter Parry was the highlight - excellent!

Confirmation regarding skills gap in the industry.

Presentation by Peter Parry.

Interaction with peers and networking.

Clarification made with the audience between digitalization and digitization.

Solving E&P problems with digitalisation November 19, 2018, London, Attendees

Hugh Ebbutt, Director, A T Kearney

Tim Papworth, General Manager Armenia, AGAPE Armenia

Peter Parry, Partner, Bain & Company

Simon Berkeley, Director, Berkeley Associates

David Sendra, Associate Consultant, BlackRockQI

Robert Kennedy, Commercial Director, Caithness Petroleum Limited

Will Jeffery, Senior Offshore Interpreter, CGG

James Foulkes, Client Support, CGG

Chris Hough, Subsurface Data Coordinator, Chevron

John Glass, MD, Cloverfield Consulting Ltd

Diwin Amarasinghe, Geophysical Specialist, Consultant

Micky Allen, Consultant

Richard Walker, Consultant Geophysicist, Cornhill Economics Ltd

Christopher Frost, Lead technical Analyst, DataCo

Michael Stewart, Operation Manager, DataCo Ltd

Dave Wallis, Senior Advisor, Energistics

Nnamdi Anyadike, London office, Energy Correspondent

Martin Blindheim, Managing Director, Energy Growth Partners Limited

James Gray, Data Administrator, Evaluate Energy

Jonathan Moore, Product Manager, Evaluate Energy Ltd.

Avinga Pallangyo, Events Manager, Finding Petroleum

Karl Jeffery, Editor, Finding Petroleum / Digital Energy Journal

Dave Camden, IM Consultant, Flare Solutions

Glenn Mansfield, Director, Flare Solutions Limited

Simon Cushing, Research Director, Gartner

Evnika Polovinkina, Managing Director, GBC Ltd

Mike Simmons, Technology Fellow (Geosciences), Halliburton/Neftex

Waclaw Jakubowicz, Managing Director, Hampton Data Services

Arun Samy, Technology Consultant, Hatch

Lawrence Jackson, Senior Account Executive, IHS

Neil Simons, Independent Consultant

Nick Steel, Independent Consultant

Simon Kendall, CEO, Interica

Christian Bukovics, Independent Director, JKX Oil&Gas Plc

Peter Allen, Consultant, Layla Resources

Alan Smith, Director, Luchelan Limited

Paul Spencer, Senior Production & Seismic Data Manager, Lynx Information Systems Ltd

Christian Fenwick, CEO, MapStand

Abi Mirkhani, COO, OPG Supply

Emmanuel Pettinotti, Geoscientist - Tertiary Team, Ophir

Daniel Buckingham, International Finance Broker, Pronto Business Funding

Ewan Makepeace, President, PT Jawasoft

Patrick Taylor, Director, RISC (UK) Limited

Hector Williams, Senior Research Fellow, Robert Gordon University

David Jackson, Principal Geologist, Shearwater Geoservices

Tom Martin, Director, Shikra Consulting

Stephen Ward, Head of VMM O&G, Siemens PLC

Andy Harris, SpectrumGeo

Duncan Irving, Practice Partner, Oil & Gas, Teradata

Peter Roberts, Business Development Manager, Tessella

Warrick Cooke, Consultant, Tessella

Jill Lewis, Managing Director, Troika International Ltd

Ugur Algan, Director, Volantice Ltd.

Robert Hayes, Subsea Support Manager, Wood Group