Data managers need to use judgement

Oil and gas data managers rarely have the resources to keep the company data perfect, so they need to use their judgement on where their time adds the most value, said data management consultant Philip Lesslar, formerly with PETRONAS

In all his years of working in data management in oil majors and National Oil Companies, Philip Lesslar has never seen a team who can meet all the organisational demands, because it would take more staff members and skills than the company is ever willing to pay for. So data managers need to learn to prioritise their work, spending their time on projects with the biggest impact.

"You can't do everything with the time you have - or the resources you have," he said.

Rather than ask how good your data can possibly be, it may be better to look at the minimum standard for data to be workable, he suggested.

Or it might help to look at the specific problem areas, where company staff do not have quality data they need to make decisions. This could be due to a lack of governance structure around data, a lack of standardised workflows governing how data is created and stored, or a lack of an accountability structure for ensuring quality and availability of data.

Mr Lesslar is formerly principal consultant, Technical Assurance, Compliance & Technical Assurance, Group Technical Data, Project Delivery and Technology, with PETRONAS Exploration and Production. He was speaking at the Digital Energy Journal forum in KL in October, "Improving the Digital Platforms - Data Management & Quality."

Not well understood

It doesn't help that data management is not well understood in the wider oil and gas industry, with wide misunderstanding on what data managers do, how much work is involved, and why it is difficult to automate.

Many oil and gas people see data managers like a company librarian. But if data management really was simple, the problems probably would have been solved by now, particularly considering how many people with PhDs have been working on it, Mr Lesslar said.

Senior managers and others continually underestimate the complexity of the task. It is common to hear senior managers make statements like "we need to keep all the data," or "transfer only the data you need", "I want it fully integrated", "we must have quality data".

Philip Lesslar, data management consultant

They say this and then they leave the room. But a data manager can't deliver this without knowing what "all the data", "the data you need", "fully integrated", and "quality data" actually mean in detail, he said.

Sometimes people profess an overconfidence in automation tools to solve the problem, saying that data can be integrated "just by writing some code, he said.

Perhaps data managers should get better at being able to demonstrate the value they create.

They can show that good data management should lead to better decision making, which leads to profit – but then it is hard to assign the profit directly to the data management.

Conversely, if the company makes losses, it could well be due to a bad decision, which could be due to poor data, but it is hard to know for sure.

Activities and data types

Mr Lesslar has a list of 28 different activities a data manager might do, with increasing complexity, from handling tapes and media, through data loading, requirements definition and standards implementation, up to data mapping, quality metrics, data science and machine learning projects.

At the bottom of the list are tasks which can be done as a "data service" from a centralised location, which are highly repetitive – including handling media, scanning, converting data, and basic data cleaning.

Tasks towards the top of the list require much more specialist knowledge.

The full list is available in his slide pack online (see link at end of article).

The list of tasks could be seen as a competency framework, where people gradually work their way up the list so they can do more complex tasks. Data managers should also see it as their responsibility to develop competence to get up the list, rather than endlessly doing data loading.

There are over 100 different upstream data types, across geology and seismic, interpretation and compilations, petroleum engineering, and drilling / production operations (which Mr Lesslar also lists in his slides).

Serving the maximum

One approach is to make sure the work you are doing is helping the maximum number of people.

Perhaps 10 per cent of the subsurface professionals in a company are doing detailed petroleum engineering or production geology studies, such as borehole imaging. These people "will find ways to take care of themselves."

But 90 per cent of staff are doing basic geological interpretation. To do this, they typically need 8 basic well logs – gamma ray, Sonic, Density, Neutron, Resistivity (S, M, D), and Caliper. So it makes sense to focus on making sure this data is available.

Selecting priority wells

A company with 10,000 wells might have 200 to 300 "of active interest" at any time. So the data management work can have more impact if it just focusses on improving data quality for these wells.

You need a process to periodically update the list of "wells of active interest", and an "enterprise dashboard" to track your progress improving the data for them.

A well header can include 40 different attributes. The well ID is a mandatory field, and usually system generated. Therefore it will

always be filled. But other attributes such as "date reached total depth" are also important but whether they are entered is a "hit or miss thing," he said.

So you could improve the basic well data by aiming to populate all of these important attributes, and check the data to ensure they are correct.

You could also make sure that the well logs are easy to find, perhaps by putting all the quality controlled well logs in a separate area of the well logs library, all given a quality stamp. Users can often get confused with all of the well logs in the database, many of which are not relevant to them and they have to make sense of all that. "I see that all the time," he said.

Digital Energy Journal's KL conference in October

Because of the diverse nature of data quality improvements, i.e. different data area focus in different wells, it is just too easy to lose track of what has been done over time. Therefore the enterprise dashboard mentioned above, that keeps track of the quality checked data, is really important. This is called the Quality Data Inventory and is essential to ensure that as you always know what has been quality checked and what still needs to be done.

With that dashboard in place, the business will help you define priority wells but you must have that process defined that will ensure you cover all business units and their priorities.

A slide is included in the presentation to explain how this is all optimized.

"All the time you know exactly where you are in the process. Every step of the way you are contributing to better business performance."

You can watch Mr Lesslar's talk on video and download slides at **www.d-e-j.com/video/1911.aspx**

Using graph databases in E&P

Graph databases, which use abstract data structures rather than rigid boxed data structures, could prove very useful in E&P, particularly working with complex data sets. Michelle Lim from PETRONAS explained

Graph databases, which store data in a loose structure rather than in rigid tables (as the usual relational databases do), could prove very helpful in exploration and production, particularly working with the large complex data sets, said Michelle Lim, from the Digital Innovation, Strategy and Architecture department of PETRONAS.

She was speaking at the Digital Energy Journal conference in KL in October, "Workforce of the Future: Improving Data Analytics & Knowledge Management."

The word "graph" means "graphical", or arranged in free space (not a graph of the sort where x changes with y).

Data in a graph database is stored as "nodes", and the database understands how the various nodes connect together as "edges".

The more common relational databases store data in tables, like data in a spreadsheet. They are much more rigid.

The development of graph theory is attributed to Euler, a mathematician, in 1736, who wanted a mathematical way to work out whether it would be possible to cross each of the 7 bridges in the town of Königsberg (now Kalingrad, Russia). He drew a map of the bridges and land connecting them as edges and nodes.

Michelle Lim, PETRONAS

Today, graph theory is used by Google, to show how different pieces of information are related. This enables Google to present a wide range of different information related to the subject you are searching for (shown on the right hand side). For example if you search for Tom Cruise, you will be given different pieces of biographical information, quotes, movies, social media pages, and pages of former partners.

Graph databases are also used by Google Maps directions, to keep track of the different ways for getting from one point to another.

Facebook is also a large graph, where it keeps tracks of all kinds of relationships between people and things. Facebook is looking at ways to develop its social graph for workplaces, to help find the right people within the organisation.

The computer system can understand data similar to how a person would – for example there are certain pieces of data which could be connected to a person, such as their date of birth and their first school. Every company office should have an address

NASA

Graph databases were used to index documents by the North American Space Agency, showing how the different documents relate, she said. This can prove a much easier way to search for a document than by keyword.

In one project, engineers wanted to develop an "up-righting" mechanism for a spacecraft, and they thought it was very likely someone had done this in NASA before, and wanted to see how they did it.

If they were searching through documents via keyword search, it could still take months to find the information, with tens of different terms which might be present in a useful document.

But with the documents connected in a graph, showing how they linked together, it was possible to find the relevant information in 4 hours, a significant cost saving from time saved.

The E&P industry has a similar challenge, with about 100 different data types. "The more information we have, the harder it is to find information, this is our dilemma," she said. "When we have massive amount of data, they are sitting across disparate databases, that makes things even harder."

E&P projects

Graph databases were used in the E&P industry to get a better understanding of well data and oilfields. The data could be visualised with green nodes for the oilfield, blue nodes for wells and purple nodes for the oilfield.

The data can be visualised with wells linked to the oilfield they are in, and wellbores linked to the wells they are in. This immediately provides a different and unusual way to visualise the company's wells, showing which fields are the largest.

This work can also show up problems or faulty data, for example if you have a wellbore which is not related to any well, or an oilfield which does not have any wells in it.

Graph databases could also be used to understand the software applications which the company is using, showing the types of data they work with. The data types are shown as green nodes and the applications are yellow nodes, and the data types are connected with the applications they work with lines.

This will generate a visualisation showing the most important data types, because they are used by many different applications. It can also help you spot where you two applications might be doing the same task because they use the same data types.

There are thousands of different software applications being used across PETRONAS and big savings if some of them can be taken out of use.

The technology might also be helpful in project management, when many different data types need to be bought together to better evaluate options for developing a field.

Graph data can also be useful for more complex data search systems (sometimes known as 'cognitive search') because it makes it easier for a computer to pull data from different data storage systems, and the graph database can better understand how the data connects together.

As the amount of data grows, and there are more different data storage systems, it is getting harder and harder for companies to get the right data together.

The data can be stored in different places – each 'node' in the database can be a separate data store.

Building it from documents

The work of building a graph database for documents would involve taking some of the main topics from the documents, and then mapping how the documents connect together.

Ultimately it could include millions of documents (as nodes) and connections between them (as edges). But you don't have to include all the documents to make a useful graph database – you can start small, say with 100 documents, and then build from there.

It usually requires domain expertise to know what fits together with what. "Humans are always required," she said.

Subsurface data analytics is changing

The way oil and gas companies are approaching subsurface data analytics is changing, with more focus on the quality of data and data preparation, and companies often building their own software rather than buying it, said Jane McConnell of Teradata

Oil companies are putting more focus on the quality of data, and getting data preparation right, before they do data analytics, said Jane McConnell, Practice Partner - Oil and Gas at Teradata, speaking at the Digital Energy Journal Kuala Lumpur conference in October, "Improving the Digital Platforms - Data Management & Quality."

Companies are often building their own analytics tools, rather than buying software to do it, so they can put data together any way they want. "The way people are interacting with this kind of data is changing," she said. "It is not just fixed workflows."

Important 'rules' for getting analytics right are to get the right people, the right platform, work with 'good enough' data management, to be 'agile', and to get business buy-in, she said. It is important to understand why subsurface analytics is different in many ways to analytics in other industries.

Good quality data

Analytics needs good quality data, which means that the data needs to be well managed.

This probably means that the industry needs to move from manual data management to more automated methods, she said.

The industry has long employed a large army of data managers moving files to wherever they are needed and do necessary conversions along the way, for example to export seismic SEG-Y data into Landmark interpretation software.

But along with this way of working comes a habit of fixing problems with the data just before the data is needed, rather than fixing problems with the data stores behind it, she said. Also the manual methods can be error prone.

For analytics to work, the systems need to be able to retrieve data from the company data stores in good quality.

Jane McConnell of Teradata

Teradata has seen many data science projects in oil companies where the analytics was well underway before someone realised a major problem with the underlying data, such as half of the data is in one unit of measure and half is in another, she said. As an example, "we had people working with weather data - half of the data had the wind speed in metres per second and the other half was knots,"

So people working with analytics end up spending endless time trying to resolve problems or fill gaps in the core data stores.

It doesn't help that the people who originally created the data, and understand it, are often not available to help.

Data lake

One desired end goal is described as a 'data lake', where all the data which analytics systems might need is readily accessible.

A data lake is not a physical data store, but more an architecture, where all of the data stores are available to the analytics systems, she said.

Many people have got the wrong idea about a data lake, thinking that if they just copied their data into a single file store, "as if by magic good things would happen," she said.

For this reason, some people are turning away from the term data lake, using the term "discovery lake" instead.

The analysis company Gartner has described three types of data lake. One is the 'inflow data lake', where you bring in data from multiple sources together to one place, such as a dashboard. The second is the 'outflow data lake' where one set of data serves multiple different applications. The third type is a "basic data lake" which is a starting point, with some controlled data management.

Data preparation

There is never enough time to get perfect data, so it is useful to define the minimum quality data which is acceptable for what you want to do with it, rather than try to get it perfect.

In this sense, subsurface data work is different to financial data, and most traditional data management work, where everything has to be absolutely correct at any point. Subsurface data work is more experimental.

Most successful data science projects just focus on one or two specific areas, rather than cover the whole company, she said. You may want to build a data lake for trying to solve a specific problem, and make sure the data is just good enough for that.

However, as the industry sees more successful data science projects, it is probable that it will want to do more and more of it. This will increase the need to bring in structured data governance processes. It will also increase the complexity of the data preparation work, so create more room for error.

Master Data Management (MDM) means setting up processes, governance and standards for managing the critical data of an organisation, making it available from a single point of reference.

"I can't think of any oil companies that do MDM beautifully," she said.

Spending time on MDM, across the company, is a good way to improve data management and to be better organised.

Build not buy

Until now, the upstream oil and gas industry has mainly worked with purchased software packages, which include data management, visualisation and analytics tools in the package. Data preparation work has mainly been in accordance with the requirements of the package. This could be called "buy not build".

Part of the reason for this is that corporate IT departments typically did not have much understanding of the petrotechnical and engineering domains, so it made sense for oil companies to use software developed by oil and gas service companies, such as Landmark and Schlumberger.

But this approach is not best suited to the analytics era, where data scientists want to put together different types of data in new ways to gather new insights from them. The data stores within software packages are usually in formats which only that software can understand.

It would probably be better to take a "build not buy" approach, making analytics tools as you go along, and building tools to get data to them in the right format. Data engineers can write code to run data pipelines, including tools to split files, move data and run data quality checks.

There is an open source project to develop software to automate the flow of data between software systems, which Teradata is involved in, called Kylo. Kylo builds on Apache NiFi. NiFi is short for Niagara Files, a previous name of the software, when it was developed by the US National Security Agency.

Kylo can be used to define the jobs you want to do with data such as split or check files, and schedule jobs and check for problems.

Many oil and gas companies have data stored in a range of old file formats, which they think can only be read with special software. But there is a surprising amount of open source software tools and scripts available which can work with old file formats, she said.

The important point is that "we need to move to a place where data is not held away from us by the software vendors and applications companies," she said.

This philosophy was adopted by the team setting up the "Diskos" National Data Repository in Norway, where there was a view that data could only be stored in non-proprietary formats.

Why subsurface is different

Subsurface data is different in many ways to data from most other industries, and it is important to understand these differences if you want to do data science with it, she said.

Subsurface data analytics includes a lot of measurement data, which is not found in other industry sectors. There can be very complex jargon and data structures.

Petrotechnical and engineering software systems are usually built around the specific needs of the subsurface domain. In this sense, they are different to oil and gas business IT, which is similar to business IT in other industries, she said.

So oil and gas business departments have long been doing analytics with the same software that other business departments use, such as Tibco Spotfire, but the petrotechnical world has been limited to what it can do within the subsurface software environment. And it has not been very easy to do analytics which involve bringing petrotechnical and business data together.

Another issue is that much subsurface data management work evolved out of work to do records management, looking after physical items such as tapes, printed well logs, fluids, seismic tape. The culture is around making sure the original data doesn't get lost, rather than finding ways to move forward, she said.

Rules for analytics

When doing subsurface data analytics, the first rule can be to get the right people. The best people are described as "T shaped," having both depth (being very good at one narrow aspect of E&P) and breadth (understanding how it all fits together). For data science projects, you want data scientists who have in-depth data science skills, but also who understand

the broader oil and gas domain. And you will need subject matter experts who have in-depth skills for their domain, but also understand the broader analytics approaches.

The second rule is to work on the right platform (software system). The E&P industry typically works with linear workflows, where data is worked on with one application in one department, then sent on to another application in another department, and these methods have evolved over time.

But this means that some data types have never been put together, because the traditional apps don't have a way to do it. Trying out new ways to put data together is usually a big part of analytics work.

Analytics also often involves looking deeply within data to see if there is something worth looking at further, which is not something which can be done easily if the data can only be accessed via an application.

The third rule is to work around "good enough" data management – doing the minimum amount of work to be able to answer the business question you want to answer. This might mean storing data so you can just pull out the piece you want, such as individual seismic traces, or well log data just for a specific depth level.

It helps if data is "profiled" so people can get an idea of what it is, without having to load it into the right software system to understand it.

The fourth rule is to be "agile", or focus narrowly on what the goal requires. The industry can so easily get stuck into "waterfall" rigid step by step processes which take years, rather than going as fast as possible to answer the specific question.

"If you have a business request for a piece of work, do that piece of work, make sure the value is delivered, don't turn it into a 10 year project," she said. Small projects can be better – people working quickly to see if they can achieve some specific outcome which is useful for the business.

The fifth role is to get business buy-in. If you want to make changes across silos of the business, you need business support at level which the various departments will both listen to, which might mean "C" level. Otherwise you can't escape out of any silo. One idea is to have a C level "chief data officer" who guides the company on things to stop doing or start doing.

The chief data officer might also try to stop people using software tools which are very difficult for someone else to work with, such as Excel and PowerPoint. They can also ensure continued governance on the data and continued data quality improvement. Dashboards can be a good way to drive data quality.

Companies are increasingly forming "asset teams" where people with different disciplines work together on business problems – rather than in the old days where one department works on data and throws it over a wall to the next one – geophysicists do their seismic interpretation, reservoir engineers do their simulation.

digital energy journal

Paradigm k – a new cloud-based system for production engineers

Oil and gas E&P software company Paradigm has developed a cloud-based system for production engineers that provides well surveillance data, reservoir simulations, and online collaboration capabilities.

Oil and gas E&P software company Paradigm has developed a cloud-based software solution for production engineers called Paradigm k, to help them perform reservoir simulations and production surveillance analysis.

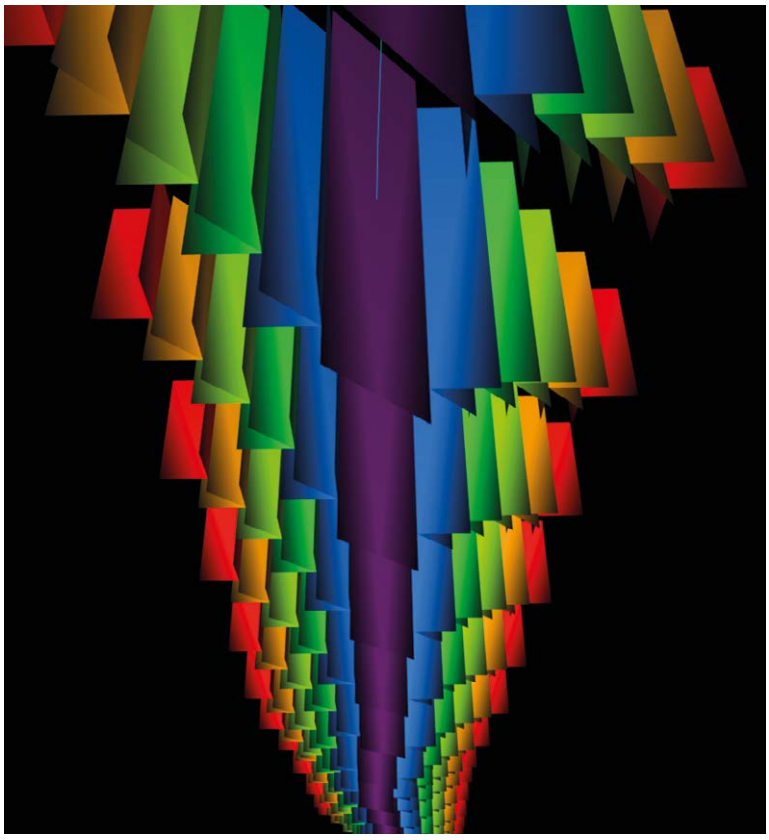
"The task of production engineers is to maintain production targets. Historically this has meant mainly surveillance, seeing what production currently looks like. Access to reservoir simulations will give them a better understanding of why changes in production rates are happening," says Indy Chakrabarti, senior vice president of Product Management at Paradigm. "It is a merger of subsurface and surface workflows coming together."

The system can be used by production engineers to test different plans for in-fill wells, and predict how much oil they might produce. They can investigate reasons why a well is not achieving its target production, or predict what might happen if you change the choke size or do an artificial lift.

The software can 'ingest' surveillance data from well flowmeters and sensors. This data can be used to update the reservoir model.

The software also supports collaboration, making it possible to share what you are doing at each well with your colleagues, as well as the results of those activities.

Paradigm sees this evolving into a knowledge base around wells. For example, someone might



The Paradigm K reservoir simulator

Understanding digital and security competency

The Malaysian Oil and Gas Services Council (MOGSC) has a project to better understand the competencies needed to do oil and gas jobs, focussing initially on ICT and security skills.

The Malaysian Oil and Gas Services Council (MOGSC) Competency and Training Working Group (CTWG) has a project to try to better understand the competencies required to do different oil and gas jobs – and which competencies are useful for a variety of different jobs. The project is focussing initially on ICT and security skills.

The study is led by Dr Jeffrey Bannister, a consultant with Orbitage. Dr Bannister also works extensively with the telecom industry. Telecoms are similar to oil and gas, in that it is very technical but also in many cases very traditional, and gone through a number of major transformations, including the current 'digital transformation', he said.

He was speaking at the Digital Energy Journal forum in KL in October, "Workforce of the Future: Improving Data Analytics & Knowledge Management".

Challenges

Companies often find that they need people with digital skills very urgently, so they rush to fill gaps rather than developing longer term plans.

Companies often find that graduates do not have the right digital skill sets, perhaps because there is not very good communication between industry and academia about what skills are needed, or graduates have theoretical knowledge but not practical knowledge. Companies complain that new graduates are ready for training, but not ready for work.

But on the other hand, companies are often too narrow when they specify what they need, saying they need someone who can do a specific thing, not the broader competencies.

It can make sense to define a set of competencies as the minimum someone would need to be able to do a job, rather than describe what the person currently doing the job can do, he said.

Often jobs are split (so 3 or 4 people doing one job) or combined (2 or 3 jobs become one), as the amount of work goes up and down. Sometimes there are competencies which companies sometimes need, but they say they don't actually need it right now.

A "competency" can be defined as a mix of knowledge, skills, behaviour and attitudes required to successfully perform a task, he said.

Roadmaps

The MOGSC project aims to build "roadmaps" or competency frameworks for organisations to follow, so they have a path to improve their overall organisational competency. Improving competency means understanding what your staff can currently do, and so where the gaps are, and if they are best filled by recruiting people who have those competencies, or by training existing staff.

Because there are so many different jobs in the oil and gas industry, the project team are aiming to define the digital skills used in a variety of different roles, rather than specific competencies for specific job roles.

This will also make it easier for staff to transfer from one role to another, because the organisation will understand that the skill they have can be used in many different roles. "We can say, 60 per cent of the skillset you have here is applicable to this role here, you need to top up this part."

It should help avoid people getting into a situation where people's skills are thought to be only suitable for the specific job that they are in. This means that people can't progress anywhere unless their boss leaves. Sometimes people in this situation just get fed up and leave the company, he said.

Once competencies have been defined, they can be used by many different companies – so it makes sense that the work of defining them is done by an industry association, rather than individual companies.

Assessing competencies

Companies want to be able to measure how well certain staff members have a certain competency – which means that it cannot be defined too vaguely.

Although if it is defined too narrowly, or with a long list of different competencies required to do a job, it can be very hard to find someone who has them all.

Assessing competency can be much easier for practical skills, Dr Bannister said. Take welding for example. A welder has a certificate which shows that they have had rigorous training and assessment in welding. He is assessed on every welding task he does. If the weld is good, the test is passed.

Dr Jeffrey Bannister, Consulatant with Orbitage

But more academic and theoretical skills are typically tested in an exam, or online multiple choice test, and only a certain number of correct answers is needed for a 'pass'.

Sometimes course trainers are explicit in telling students, they are not there to be trained in skills they will need in the workplace, they are there to learn how to pass the exam.

It may be better to test ICT skills in a similar way to how the welder is assessed, such as by asking people to configure a firewall or demonstrate that they can analyse data.

Learning from telecoms and IT

The telecom and IT industries have already done a great deal of work defining competencies in digital and security skills, which is something oil and gas might be able to make use of, he said.

One Malaysian organisation, CCPS, has developed a model showing different roles as a tower of blocks. Where the same skill is used for different roles, the block stretches across different towers. Each box has a description of what performance outcomes someone should be able to achieve to demonstrate they have competency.

This framework sets a common language between HR, learning and development, and technical teams. It can make it easier to identify which skills are easy to find among new recruits, and where it makes sense to train the existing workforce. It can also help communications between HR and technical departments, if technical people can just point to the skills they need as one of the blocks.

MOSGC has taken on three areas of this framework into the oil and gas industry, covering IT (databases, software development and analytics), IT security, and IT data communications and networking. Each area has the same fundamental, transferrable skills, of basic networking and basic computer operating.

MOSGC is circulating it around the oil and gas industry to get feedback on how it can be adapted for the industry and where bigger changes are needed and where the gaps are, he said.

HR / technical communications

One challenge is the quality of communications between technical staff (who are often working on today's problem and finding they don't have people to do the necessary tasks), and HR people (who have the role of finding people who can do the tasks and help them develop skills).

Typically, the HR people will not understand the technical domain. They ask technical people what skills they need. The technical people come back with a list of skills which HR people do not understand, which they have to hammer into a plan or a training course.

The training course can often mean someone in HR just passing on the list to a 'usual suspect' training provider and ask them to make a 2 day program. The end result is often unsatisfactory.

One common problem is when the company gets a new HR manager, who wants to improve the way the company manages skill developments. This person is typically not technical, so feels more comfortable with the so-called "soft skills" training, such as leadership and communications, so this is tackled first. But the technical part does not get tackled well at all.

HR people ask technical supervisors to do periodic 'appraisals' of their staff, but the value of these is questionable. "I was with the CTO of a large organisation recently, and he said, let's be honest, by the time I've got to the 8th or 9th appraisal that morning, number 10 is not going to have much thought put into it. The further in there we go, we get less effective."

You can watch Dr Bannister's talk on video and download slides at
www.d-e-j.com/video/1907.aspx

CISCO – making your "internet of things" easier to manage

Computer networking giant CISCO has developed an 'internet of things' platform called Kinetic, to make it easier for companies to manage complex networks of sensors and the data they generate

CISCO, a company best known for computer networking equipment, has developed a product called Kinetic for companies to manage complex networks of sensors and other 'internet of things' (IoT) devices.

The service is most beneficial for companies which have thousands of different devices, and want a way to reduce the headache of setting the devices up, managing the policy of what data they want and where they want it, running appropriate software on the devices, and then actually moving the data.

The service is provided as a mixture of hardware (Cisco routers) and software running on cloud services.

You can immediately set the whole system to a default setting, where your sensors and other devices provide standard data to cloud hosted software, ready to be fed into your various software applications. You can then use this default setting as a starting point for configuring the system to do exactly what you want.

For the oil and gas industry, Cisco provides a starter kit, as a "default blueprint" to get you going. It includes a default way to extract data, and cloud hosted applications which can receive the data on a daily, weekly or monthly data.

The oil and gas industry uses many different "internet of things" devices, including acoustic sensors, temperature sensors, fibre optics, gas leakage monitoring and equipment health monitoring.

Apart from oil and gas, the system is used in city management (including lighting and parking), manufacturing (monitoring machines, energy, inventory and deliveries), transport (including traffic lights and 'connected car), and retail.

Theresa Bui, director of IoT strategy with CISCO

The service is offered following Cisco's acquisition of Jasper, a company which makes a cloud based 'internet of things' service platform, in March 2016. It is used by over 16,000 companies.

Working with well data

One oil and gas industry customer uses the system to gather and work with well data.

It has thousands of wells worldwide, all fitted with different sensors, generating data in different formats, and with different "application programming interfaces" (APIs). This meant that gathering and managing the data involved a great deal of manual work. It was very difficult to build a centralised system for understanding the whole 'fleet' of wells, and do any data analytics.

The company wanted to get visibility in near real time of what was happening with the wells, and wanted it all automated, so it would not require the data to be formatted by an engineering team before it could be used.

The oil company had four different business units who wanted to work with the well data and do different kinds of analysis on it.

They wanted statistics for individual wells, and aggregated information about multiple wells, and the ability to do big data analytics.

CISCO installed a new router on each of the customer's sites, which connected upstream to the various sensors on the rigs, and downstream to the cloud hosted software. From there the data could be distributed to the software systems of the business units who wanted to work with it.

Setting up the system

When setting up your IoT system, you need to determine what data you want to receive from all of your sensors, and where this data should go. This is called creating data "policies".

For example, if you have 32 sensors on a well, you might want sensors 1-10 to send data to SAP hourly, and send data from another sensor once a month.

A company might want its business partners to see some of the data but control exactly what they can see.

It is much easier to set up a system if your sensors are more standardised, generating data using the same data model, says Theresa Bui, Director of IoT Strategy, CISCO.

Various options are available for connectivity, including cellular, and a new satellite com-

Improving well head data at Brunei Shell

Nurhamizam Yussop of Brunei Shell Petroleum talked about how the company is improving the way it manages data about well heads, speaking at our Kuala Lumpur conference in October

Brunei Shell recently embarked on a project to improve data management of its well heads – including trying to get correct data of all well head configurations.

Nurhamizan Yussop, wells and production technology data technician with Brunei Shell Petroleum, explained how it is being done, with a talk at the Digital Energy Journal Kuala Lumpur conference in October.

Mr Nurhamizan Yussop works in the technical data management team, of Brunei Shell, looking after the wells and production technology discipline's data.

The aim was to get a "single source of truth" version of well data, which could be made available to all staff, with all verified well head reports on one platform, with the "engineers Data Model" software by Landmark being chosen.

The company standard is that all mechanical data for the well head / Christmas tree should be recorded in the EDM.

The data has a vertical component (the parts of the well head handling flow upwards from the well, including well hangers), and a horizontal component, the assembly to take flow to the production tubing, including annulus valves.

Before the project, data about well heads was being stored in various formats in the company, including electronic well files and physical well files. Data is held by both the production and wells teams.

Nurhamizam Yussop of Brunei Shell Petroleum

Data has been captured in the database at different stages of the project, from engineering design to execution.

The well head is a structure on the surface which provide the suspension point and pressure seals for the casing strings, and also where the production tubing is connected to. There are assemblies of valves sitting on the wellhead to control flow. These are called Christmas tree.

If the company has incorrect data about the specifications of the wellhead, it can lead to incorrect calculations about Maximum Allowable Annulus Surface pressure (MAASP), Operating envelopes production volumes, time consuming work to correct errors, and frustration.

Brunei is about the size of Singapore, covered with tropical rainforest. Oil and gas is the main export.

The oil and gas assets are split into 3 regions – onshore Darat, offshore East and offshore West. There are over 4500 reservoirs and over 244 offshore structures, manned and unmanned. There are about 1600 producing wells, including water injection. Total production is 350,000 boepd. There are over 3,000 staff.

It is the main supplier of domestic fuel to Brunei – and also gas is liquefied for export.

The project

The company's data clean-up project began at the end of 2016, stating that specific information needed to be recorded for all new wells. Then the work to improve the data from the existing wells began.

The project team aimed to work closely with stakeholders, including the well engineering department and production technology department, to understand where the current data systems can be improved, and what process should be followed to improve it.

The project team also made use of a Shell online forum called Shell International Global Network, where Shell employees share best practises and answer questions. They found that Shell Canada had undertaken a similar project, and were able to discuss with the people involved.

A "standard operating procedure" was created, showing how new data should be created and entered into the EDM software. This procedure was put through a test.

The fields on the form are colour coded, where yellow means they are mandatory for the business, and white is "good to have".

A data quality team was established, with a role of checking all new data entered into the system, whether it is correct and whether it is entered in accordance with the standard. Once approved, data is given a 'stamp', including with the time it was approved. This means that anyone working with the data later can be sure that they have the latest approved data.

There may be a need to check that the data is in alignment with what has actually been installed on site, for example by asking the drilling supervisor to confirm the data is correct.

There is also a traffic light system, where data is shown as "green" if it is approved, "amber" if it is in the approval stage, and "red" if it is in the compilation stage.

The colour might change if there is a change to the data's status, for example data is updated during maintenance work, but the update has not yet been approved.

The company normally runs automated checks through all of the data at the end of every month, checking for data completeness.

Mr Nurhamizan said that they had completed a previous project of mechanical backloading for 1773 wells into EDM that took about 4.5 years to complete. This time, with the help of the best practices, they will aim to complete wellhead backloading project into EDM in 2 years.

Ultimately, Brunei Shell will have all of its well data verified, and available immediately to staff, from a single database. It will lead to better calculations, better and faster decision making, and problems due to bad data being avoided.

Brunei Shell – Spotfire for well performance review

Brunei Shell is using Spotfire software for reviewing the performance of its wells. Jamree Ismail, production geoscience technical data consultant with Brunei Shell explained how it worked

Oil companies need to periodically evaluate the performance of their wells, and just gathering the necessary data can take three weeks, said Jamree Ismail, production geoscience technical data consultant with Brunei Shell Petroleum.

Data needs to be gathered from many different databases and applications, including data about flow rates and pressures, data about the reservoir and what it might be able to deliver through the well, and nearby drilling opportunities which might be viable.

The data is sometimes only available in spreadsheets or PowerPoint slides, rather than from a central corporate database. Sometimes the co-ordinate reference systems on geographical data do not match. Some data sources are more trustworthy than others, and some data is not available at all. "That's why it takes 3 weeks to do it," he said.

Software like Spotfire can be used to connect with a number of different databases and data stores at once, including subsurface modelling software like Petrel, and surface data. It can be programmed to put all of the relevant results together in one visualisation, called a 'composite map', he said.

For example, the basic geographical map might come from Petrel, the production data from a database, the reservoir pressure gradient (change) and well location from another system. Data about past production of wells can be made available by 'drilling down' from the map.

It connects directly to the corporate database, there is no Excel "staging layer".

The images on Spotfire can then be used to make PowerPoint slides,
The images can also be taken out of Spotfire and included on PowerPoint slides. They can be used as a basis for calculations.

Spotfire proves to be a useful tool for managing data quality, when the visualisation makes it clear that some of the data does not fit with the trend of the rest.

Just by being able to work with data directly in the corporate database, can save 120 man hours on one project – and there might be 50 projects going on, he said. Multiplying this to show the total man hours directly saved really helps quantify the value of it, he said.

Jamree Ismail, production geoscience technical data consultant with Brunei Shell

The project has been presented to the Brunei Shell managing director.

Spotfire is not new software, but now there is more support in the company for the idea of using these tools in a sustainable way. "We're trying to put this as a template for the enterprise that everybody can use," he said.

"We used Spotfire not only from a data quality perspective but we took this tool and look at ways that can change our process to be more lean and competitive."

Measuring quality of national oil and gas data

Philip Lesslar, data management consultant, presented a a summary of the activities of the Energistics National Data Repository (NDR) workgroup on data quality metrics, talking at our Kuala Lumpur conference

The article summarises the activities of the NDR (National Data Repository) Workgroup on Data Quality Metrics. The membership of NDR are mainly government oil and gas companies but include also IOCs (International Oil Companies) and service providers, as presented by Philip Lesslar at the Digital Energy Journal Kuala Lumpur conference in October 2017.

NDR is run by Energistics. This data quality workgroup has documented their work from 2012 to 2017 into a 3 volume guide for implementing a data quality metrics program.

The lead participant is Philip Lesslar, currently an independent data management consultant, and formerly principal data management consultant with PETRONAS Exploration and Production and before that Shell International Exploration and Production. The three other participants that make up the DQ Workgroup are: Helen Stephenson, Stephenson & Associates, Ugur Algan, Volantice and Jill Lewis, Troika.

The work started in October 2012, at the NDR conference in Kuala Lumpur, with a discussion session to try to understand what kind of data quality issues organisations face and how they could be resolved, he said. The next 3 years were spent collating information and providing introductory write-ups on data quality.

The next stage of the project was to try to establish a standard set of 'business rules' which good quality data would follow. By assessing whether the data passes the rules, you get a sense of the quality of the data.

All of the people in the project team were data management professionals.

Continuous process


Philip Lesslar, Data Management Consultant

The challenge is to 'operationalise' the data quality metrics process, which requires agreed business rules to be implemented as queries in data quality tools that can be automated to fire these rules at regular intervals. With this in place, you have a continuous wave of data quality checking activity which frees data managers from all the mundane checking they

would otherwise need to do.

It is too easy for data managers to just get stuck in a routine of finding errors and correcting them, he said.
Once you have such an automated system that can run millions of checks on data, data managers just need to set up a 'service structure', constantly monitoring for exceptions, ways to fix these and also develop enhancements to the overall process to improve it.

Different problems

Every data type has its own problems, he said, including geological data and interpretations, drilling data and production data.

When planning a data management project, it can be very hard to predict how long it will take, and how many people it will need, to ensure you get the results you want.

Psychology has a big place to play in the world of data management, because it is easy to lose morale in your efforts to improve corporate data, when no-one seems to care, or has main priorities elsewhere, he said.

End goal

The end goal everybody aspires to reach is to have trusted, fit for purpose data, which is immediately ready for use in projects. This means that less time will be required to get data ready for a new project, and decisions can be made faster. There should be dashboards telling you everything you want to know about the data. "These are things we all work towards," he said.

Good national data should also make it easier for companies considering making investments to understand what the country has to offer.

A good national data repository should have a consistent set of business rules, for example saying that all data entered must have certain fields completed, and it must be in a certain structure or format.

"If you have quality, fit for business data, your business workflows will be much smoother."

Metrics

It is important to have metrics showing how much a certain piece of data meets the rules and how complete it is. Like with any other business activity, it is important to know how well you are doing.

But just having a metrics system does not necessary lead to improvement. Improvement can only be achieved by putting a holistic process in place and ensuring that it sustains. Data quality dashboards which can tell you how many errors there are in the databases, but if the meaning of the errors is not understood, correction can be difficult.

For all your different types of data, you want to know how much of it is good and what specifically is good, so you can make sure you only use good data.

The question of how good the data needs to be is a separate issue. The data needs to be fit for the purpose it is to be used for.

A good starting point is to understand that certain workflows (such as exploration, drilling), need certain data types, so the data manager should focus on making sure that good data is available.

Primary and secondary

Data can be classified into primary and secondary data. Mr Lesslar provides a slide that classifies data into the main primary and secondary categories with examples of each. "When we discuss data, we need to be clear which category we talk about as each they need to be managed differently" he said.

Primary is sometimes called "raw data", or "master data", coming directly from the sensors.

Secondary data is created from the primary data when people in the various EP disciplines do their interpretation. The number of secondary data types tends to me more than that of primary data and is harder to manage. You have data of different versions, vintages, and data updates.

The thing one sees when people do interpretation is a tendency to create different versions of a geological pick, for example. However, at the end of the project, housekeeping is generally not done and these different versions of the pick are left in the project repository. When several interpreters in the project create several versions each, the multiples become large and very confusing. But no one wants to delete them "just in case" Mr Lesslar likens it to TV programs when you see someone's house where they have been hoarding newspapers for years in piles rising to the ceiling.
"Sometimes we behave like that with data," he said.

When data was only stored on mainframe systems, managing it was a lot simpler. Now lots of data is stored in personal stores.

Data analytics and master data

There is a growing interest in data analytics, and it is worth noting that data analytics might require a higher quality of data than other uses, he said.

The basic architecture for analytics is "very simple", starting with corporate data which are usually in relational or hierarchical databases, then data being loaded into an analytics system, that usually has a data warehouse construction and a dashboard for visualising the results.

The analytics and visualisation is done directly from the data warehouse. Data is pulled from the source corporate databases using ETL (extract, transform, load) processes that usually included some form of cleansing. Additional quality checks and infill is likely to be done in the data warehouse itself to ensure the analytics and visualisation goes smoothly.

But that means that the cleaned data does not get put back into the corporate database. There is no process for doing that, and if there is a process, the obstacle is that data models are different. It is easier to ignore the problem.

As a result, corporate databases, which is supposed to be the master and cleanest source of data, actually becomes of less quality than the analytics data warehouse. When there are several such analytics implementations, the problem is compounded.

The Data Quality Metrics Guides

The NDR Data Quality Working Group have completed and published 3 guides and made these openly available on the Energistics website. The link is given below.

Part 1: Background and Case for Change. Context and justification to Management

Part 2: Business Rules Fundamentals. Data quality dimensions, key concepts around business rules, 18 data types, 241 rules

Part 3: Implementation. Metrics, dashboards, implementing rules as queries, understanding results, getting the program going

Download the NDR data quality guidelines at http://www.energistics.org/standards-portfolio/standards-download

Turning old pipeline data into an engineering model

Data consultant Piyush Pandey was recently given a project to turn a large pipeline legacy data archive from an oil major from the Middle East into a modern, integrated engineering pipeline data model. Now Director at Geologix, Piyush Pandey explained how it was done.

Data consultant Piyush Pandey was recently given a project to try to turn a large volume of old engineering data into a single engineering pipeline data model.

The data was owned by a gas pipeline company in the Middle East, which had made an acquisition.

The ultimate aim was to have a single engineering pipeline data model with all of the useful data integrated into it, and all of the data in it correct.

The engineering data model would be used to plan pipeline operations & maintenance work, including identifying where maintenance was most required, so maintenance could be done on a 'risk' or predictive basis, rather than on a fixed schedule or a reactive basis (fixing problems which have already occurred).

It also wanted to be able to assess the risk of failure of an entire asset, or group of components, rather than just understanding the risk of individual components. The complex integrity model was used to schedule the predictive network maintenance program.

In order to do this, it would be necessary to work out the reliability of every single asset, and how much life it is likely to have left, he said.

The data would also be available immediately as required, in the same way that we have electricity whenever we need it, he said.

The acquired company had different ways to work with data from the new parent company, including acquiring data in different ways and storing it in different data models & data formats.

The old data

The company had about 85,000 pages of pdfs, mainly CAD drawings of various sizes. There were documents created in long obsolete software such as Wordstar, from which it was impossible to extract data. There was also a lot of handwritten information.

The work involved optical character recognition (OCR), from scanning old documents, but also what is called "intelligent character recognition" because a large amount of "intelligence" is necessary in understanding what the documents mean, he said.

Much of the existing data was typewritten, and some of the ink had faded. This is why just using optical character recognition is not enough.

Computer rules were used to try to guess what data represented. For example if a number has a dollar sign in front you know it is a financial value. If it is talking about millions of cubic feet per day it is talking about volume of gas.

Much of the data could not be read by machine. For example, an 'approval date' - a critical piece of information - could be entered on a handstamp, which was not stamped squarely on the page.

Some data was provided as a ticked box on a form. Mr Pandey noted that in the US, people select a box with a cross, but in other countries they select a box with a tick, and use a cross to indicate 'no'.

The company was also sometimes entering multiple versions of the same form into its archive. The computer program had to be programmed to automatically give documents version numbers.

Every document was scanned twice, on different scanners, and the two versions compared to identify errors. This led to a big improvement in accuracy.

Working through the documents, a barcode was put on every paper page, so it was possible to track where all information in the computer system came from.

A rule based intelligent character recognition system was used to try to automatically extract information from scanned images and put it in databases. The system started being able to capture 71 per cent of data, but with training, could get 94 per cent of data.

Drawings were also a challenge. A drawing

Data consultant Piyush Pandey

contains two sets of data – the physical drawing and information written on top of the drawings. The drawings are all in different scales and format sizes, showing different things. To get the data in a drawing into a single integrated engineering data model, you need to have all of the drawings on the same geographical scale and with a horizontal alignment.

Mapping out the workflow

The approach taken was to try to map out the entire workflow which was followed when creating the old documents.

For example, the company might put a new document in its files every time a modification or new project was approved. Key data on this document might be the commissioning date, and date it was approved. By extracting this data you would have a 'model' for understanding how the documents fit together.

In this way, it is possible to build up a model of documents made over the lifetime of the pipeline, including testing and commissioning, up to decommissioning.

Data model

All of the data that could be extracted was compiled into a single data model, using the Pipeline Open Data Standard (PODS). The data was quality controlled before importing it into the overall data model.

The data in the model was georeferenced - including data which was not from drawings - so you can click on a point in the pipeline and extract relevant data about it.

How the Malaysian government encourages digitech

The Malaysian government has an agency, "Malaysia Digital Economy Corporation" (MDEC), and an initiative, "ASEAN Data Analytics eXchange" (ADAX) to encourage more use of digital technology within Malaysian business and government as well as by its citizens.

The Malaysian government has an agency "Malaysian Digital Economy Corporation", and an initiative "ASEAN Data Analytics eXchange," to encourage better use of digital technology within Malaysian business and government as well as by its citizens, and to develop data science skills in Malaysia, among other objectives.

The projects are not specific to the oil and gas industry, but the oil and gas industry of course is a major part of the Malaysian economy.

We heard more about MDEC and ADAX at the Digital Energy Journal conference in Kuala Lumpur in October.

Somasundaram Nagappan, Head, Strategic Partnership, Data Economy Division Malaysia Digital Economy Corporation

MDEC

The aims of the Malaysian Digital Economy Corporation (MDEC) are to attract investment into Malaysia from technology companies, to support the 'local champions' developing technology and solutions within Malaysia, and to help local companies sell outside Malaysia, said Somasundaram Nagappan, Head, Strategic Partnership, Data Economy Division Malaysia Digital Economy Corporation.

It also wants to help develop Malaysia's digital economy 'ecosystem', and help companies find skilled staff.

MDEC initially had separate units to drive adoption of big data and 'internet of things' but these have recently been merged into a new division called the 'data economy', looking at how companies can monetise their data, or get useful insights from data and make data driven decisions.

MDEC works with various parties, including institutions of higher learning, targeted students in order to create awareness of the option of being a data professional as a career. "As a career it has one of the highest growth rates for salary," Mr Nagappan said.

Malaysia's estimated needs for data professionals by 2020 has risen from 16,000 (predicted in 2016) to 20,000 (predicted in 2017). So far there are 7,000 data professionals in Malaysia who have come through various programes facilitated by MDEC, and MDEC is on track to achieve the 2020 target.

There are two pathways for training – being trained in your normal university course and reskilling while you are currently employed.

About 14 universities in Malaysia are offering data science courses/modules. Some people learn through traditional IT and computer science degrees, perhaps specialising in data science in their final year.

MDEC is working as an intermediary between academia and industry ensuring that the courses meet industry requirements.

Another path is doing data science as part of Masters or Phd program.

And a third approach is where data science is embedded into other programs. "It is not an IT thing, it is used in every industry you can think of," he said. In fact, it may be easier for people trained in other disciplines to develop data skills, than for a data professional to learn another discipline.

"There's no point in putting an IT guy in the oil and gas industry and expecting him to know everything," he said. "We'd rather have someone with the subject expertise of oil and gas, but with the data science knowledge, doing the work.

About 70 per cent of current data professionals are people who developed their skills while employed in something else, rather than learning it a university, he said. But the university programs have only recently started, so it will be a few years before the students hit the jobs market.

For people looking to re-skill with data skills, there are intensive training courses available which promise to turn you into a data scientist in two days, but which have variable quality, he said.

To help people choose good courses, MDEC has been working with industry to try to develop a training framework, with a standard set of skills someone should posses so as to be able to work as a data scientist, analyst or engineer. MDEC, through its ASEAN Data Analytics eXchange (ADAX) initiative, will be able to provide endorsements to say a training course meets the required standard.

MDEC has also invited Harvard Business School to run intensive 4 day courses in Malaysia in data science, aimed at "C level" executives. There had been 3 programs completed as of Oct 2017, and a fourth starting in early 2018. "They don't let you leave the hotel room for 4 days, [presenting] use cases after use cases," he said.

MDEC is also running "Data Star," a 2 month intensive data science program available to recent graduates in a science / technology / engineering / maths subject. After the 2 months they get a 4-6 month placement in industry, and are then promised a higher starting salary than the average for new graduates. This is an elite program only available to the "best of the best," he said.

The first cohort of 48 students started in May 2017, the second of about 70 started

in October 2017. A third will start in December 2017 of about 100. This course is provided via ADAX.

ADAX

ADAX is an initiative by MDEC to enable businesses, governments, academia and professionals to rapidly adopt Data Analytics as a tool to empower decision making and innovation. ADAX seeks to be the definitive Data Analytics Exchange Hub for knowledge, information, resources and collaboration for the ASEAN region. The main goal of ADAX is to develop the ecosystem, build a critical mass of talent pool in the big Data Analytics category and to foster collaboration amongst businesses, start-ups, academia and professionals so that Data Analytics becomes an integral part of business innovation and decision making.

It has worked with a number of oil and gas companies, including PETRONAS, discussing analytics both with specific technical departments, such as well engineering, and the ICT department. It also built the framework for MDEC's Data Star project described above.

Dr Farouk Abdullah, chief data scientist, believes that mindset and company culture is one of the most important factors in getting data analytics moving. You need a 'fail fast' culture, not a culture where people are mainly focussed on following instructions and avoiding mistakes.

One way to assess your company's "maturity" in using data is to give yourself a score from one to five on your progress in analytics, data, technology, organisation culture, talent and data strategy. All elements need to be in place to become a "data company", he said.

Companies are also reluctant to invest in data analytics staff, saying that they don't see the value. But perhaps they would be better off looking at how their entire company could improve with analytics as part of its core capabilities, he said.

Also, analytics shouldn't be used as a last resort, "we've tried everything let's get the data analytics guy" – but used proactively to help keep customers coming back, he said.

Dr Abdullah was born in Singapore but grew up and worked in Scotland, and worked as head of analytics for the Bank of Scotland Corporate, then head of analytics for Scottish and Southern Energy, then for British Sky Broadcasting, then head of analytics with Deloitte.

He started his working career doing data 'analysis'. This is now more commonly called 'analytics' or 'data science', but "it is still analysis essentially, and there's nothing wrong with Excel, he said.

Service culture

Working with data analytics can help companies and departments transition from having a messaging and reactive service culture to a much more service orientated and personalised one, he said.

Consider the HR department. In the past people would typically only make contact with HR on the point of hiring, when leaving the company, or after doing something wrong. But when HR people start using analytics, they become far more proactive, using psychometric testing and studies of attitude and core skill sets in recruiting (and abolishing CVs), and supporting staff much more during their time with the company.

If you use analytics well, you can keep your service improving – rather than just aim to keep your service above a certain level, such as a certain customer retention, he said.

In order to develop a more service oriented culture, you need a detailed understanding of your customers, something which many companies still don't have. As an illustration, ADAX worked with one Malaysian company, where Dr Abdullah asked the CEO, CTO and CIO how many customers they thought they have, and they provided answers respectively of 5m, 3m, and 4m.

You need to be able to communicate with customers and employees with a 'segmentation of one', not send bulk mailings and e-mails to massive numbers of people at once. "You're not giving the person the information they need, you are giving everyone all the information," he said.

Analytics can help you understand how your industries are changing. It can show that the 'funnel' of sales and marketing is getting shorter and shorter, where instead of having a gradual increase in engagement with a company before buying, customers typically just show up and buy.

Oil and gas

In the oil and gas industry, many companies are collecting and storing lots of data, but not doing much with it, and refusing to share it.

"It is more important to understand the application of data than data itself," he said.

However, oil companies should also recognise the data competences of their existing staff. For example, oil and gas mechanical engineers are typically very good at statistics and programming, although they do not necessarily see themselves as data professionals.

"Everybody does anlaysis," he said. "If you can open a spreadsheet you are an analyst in large organisations. They have finance analysts, sales analysis, downstream analysts."

The important point is for a company to understand where its people with data analytics skill sets are. "Find someone who is doing analysis but not calling themselves data scientists," he suggested.

Oil company people tend to be focussed on one specific problem at a time, trying to optimise something, rather than taking a holistic picture about what they are doing, he said. They say they are working on an analytics project, and when they have finished it, they will do another one.

It helps if you have a non-conformist culture in your company to do data science, because people will need to be finding answers by themselves, he said.

Also, you need a solid data management foundation. "I go to organisations and people say, we have some data, have a look at it. I look at it, I saw the data quality is not there. It means we can't do anything predictive with it."