digital energy journal

Report from our Kuala Lumpur forum

A standardised structure for documents - Brunei Shell

Data on demand at PETRONAS

Condensing subsurface data to a grid

What we can do with RESQML

Monitoring compliance of wells data

How to make data science useful

December 2019 - January 2020

Can you assess cybersecurity from outside?
Integrating "wearable" safety data

Finding Petroleum

Official publication of Finding Petroleum

digital energy journal

Issue 80 December 2019 - January 2020

Digital Energy Journal
United House, North Road,
London, N7 9DP, UK
www.d-e-j.com
Tel +44 (0)208 150 5292

Editor and Publisher
Karl Jeffery
jeffery@d-e-j.com
Tel +44 208 150 5292

Production
Very Vermilion Ltd.
www.veryvermilion.co.uk

Subscriptions:
£250 for personal subscription, £795 for corporate subscription.
E-mail: subs@d-e-j.com

Cover photo - delegates from the Digital Energy Journal / Precision DM course "Data Quality Metrics Primer" in Kuala Lumpur on October 4, 2019

Opening

Data management - great ideas from Malaysia

We heard some about some great work going on with oil and gas data management in our Kuala Lumpur forum last October - and you can read about it in this issue

Poh Hean Yap, data science lead with Accenture, explained how domain expertise can be used together with data science in ways we had not heard before. She gave an example of how data science was used to identify .. that a company was taking oil sampling in the wrong place. Not something we would expect a data scientist to find, but a saving worth millions, when it was showing a company how to stop sending oil in the wrong specification to a refinery, where it was causing expensive breakdowns.

Alvin Alexander, geo technician with JX Nippon Oil and Gas Exploration, explained how you can get great E&P data from pdf files. You probably need to make - or steal - good scripting tools. He showed how you could convert online marine notices about drilling activity – "Notices to Mariners" – into a map of drilling activity planned for a certain part of the world – and gave the audience an online link so you can download his code and try it yourself. (The link is also in this magazine).

Hjh Noor Raidah Hj Basir and Nurhamizan Yussop, both from Brunei Shell, presented the interesting steps they are taking to improve how they manage technical data and documents in the company, and new systems they have implemented.

Ross Philo, CEO of Energistics, explained how companies can now move subsurface data easily from one software application to another, taking advantage of the RESQML data standard. It means subsurface work no longer needs to follow a linear workflow – and you can do analysis on a subset of the data, you don't need to work with all of it. All of this makes it possible to do much more with subsurface data.

Chew Wei Liang - Manager, Surface Operations, Group Technical Data with PETRONAS, explained how the company is developing methods which can make all the best data available to technical specialists via an online interface, and ultimately pushing data automatically into the software tools they use.

Teck Hing Wong, senior data management consultant with Sarawak Shell, explained how the company is integrating all of its best seismic data and interpretations onto a grid. It means you can achieve the geoscientist's dream of having all of your subsurface data all integrated together.

Philip Lesslar, data consultant with Precision DM (formerly with Shell and PETRONAS) gave an overview into the different methods which can be used to identify clustering and how it can be a basis of a machine learning workflow.

This may have been one of the best subsurface data management events ever held in terms of people from operators talking openly about how well it went and how well it worked. We would be very interested in doing similar style events around the world next year - if you may be interested in getting involved please let me know.

Karl Jeffery, editor

Digital Energy Journal, London

How to scrape useful data from pdfs

Imagine an oil and gas company wanting a map of all the drilling happening in a certain part of the world. The data is publicly available, but on pdf reports on a maritime authority website, for warning ships to keep clear. How can you "scrape" data from these online pdfs and convert it into data, so it can be posted on a map? Alvin Alexander showed how

Imagine an oil company executive who wants to see all the drilling activity planned for a certain part of the world over the next year, shown on a map.

Information about drilling activities is publicly available in the form of 'Notices to Mariners', documents for ship crew so they know to avoid the area. But it is provided as a number of signed online documents, not a database.

Alvin Alexander, geo-technician with the Geology & Geophysics department of JX Nippon Oil and Gas, explained how to scrape data from the pdfs and post it on a map, speaking at the Digital Energy Journal KL forum in October, "How to digitalise exploration and wells."

The same technique could be used for anything else an oil company wants to do extracting data from pdf files.

One obvious question is why the data is not issued in a map by the Malaysia Department of Marine in the first place. Mr Alexander cannot answer this but thinks it may be for legal reasons - "They need it in a pdf with the signature of the director of marine."

Although Mr Alexander's job title is "geotechnician" he sees himself more as a data manager.

He believes that it is the data manager's role to "assist and empower people" – helping them get the data they need – rather than people being slowed down in their work by having to hunt for data.

Mr Alexander shares some of the code he has developed (or taken from other places) on a GitHub page. https://github.com/elvinado/Scraping-NTM-DEJ

"You can run it tomorrow and it works," he said.

Four steps to extracting data from pdfs

The starting point with this project is hundreds of pdfs available on the Malaysia Marine Department website "Notices to Mariners", about activities which ship operators should be aware of. Not all of them are related to drilling.

For an example see http://www.marine.gov.my/jlmv4/sites/default/files/NTM1822019.pdf

The activities related to drilling can be picked out automatically, because they all have very precise geographic co-ordinates, which a computer can search for.

Normally a task to extract drilling data from these pdfs would involve downloading all the pdfs manually, searching through them for documents about drilling, and then copying the co-ordinates of the drilling, the date, and some information into a spreadsheet, which could then be loaded into geographical software.

Alvin Alexander, geo-technician with the Geology & Geophysics department of JX Nippon Oil and Gas

Mr Alexander wanted to explore whether there were automated ways to do this.

The four steps could be downloading all the pdf files, automatically searching which ones contain geographical co-ordinates (indicating that they relate to drilling), extracting the co-ordinates and outputting the data.

Downloading a website is much easier if the pages are static, rather than dynamic (with pages generated by a code) – because with dynamic pages you can't download the root code which creates the pages. But there are Python tools to automatically download all the pdfs.

The next stage is to understand the pdf itself. It looks straightforward to a person in its original format, with tables and lines, but when you use an automated text extractor, "it sometimes becomes gibberish," he said.

There are a number of free software tools available to extract text from pdfs. Some optical character recognition tools are poor. It would be useful to have machine learning based tools to recognize characters, but so far none are available free of charge.

The next stage is to extract co-ordinates from the text. Getting the co-ordinates themselves is simple, there are two co-ordinates of any point, in a very exact and fixed format. The hard part turns out to be connecting the co-ordinate with a name with a limited number of characters. And if you extract the name from the text wrongly, then a person won't know what the co-ordinate means.

The final point is to post the points on a map, with a GIS tool.

You do not need to achieve complete automation – if there is some weird data you can correct it manually – but the point is that the number of documents requiring manual editing is much lower.

The same approach could be used by oil and gas companies working with pdfs in their archive.

For example, you could search multiple documents for a mention of a well total depth.

Approach to programming

Mr Alexander taught himself programming.

He advocates trying the easiest task first. "More success builds up confidence," he said.

But there is a lot of reading and searching online involved, including when you have to look up definitions. "There's a continuous cycle of Googling," he said.

You should be aware that most programming problems have multiple solutions. If your chosen solution is not working, you can look for another one. Sometimes "tunnel vision" can be a problem, when people believe their solution is the only possible one.

Another solution to problems is to take a few minutes break, and suddenly another idea pops into your mind.

Python proves to be a "very friendly programming language," he said. "Previously – I started with Visual Basic, it is not a friendly language, there is only one way to solve any problem."

Just about anything you can see on a web browser can be downloaded, including websites with logins.

"It is quite difficult, but it can be done."

"As data managers, we need to facilitate the data to the consumer," he said. "It doesn't matter how – as long as we do it."

With enough scripting competence, you can generally write a code routine for a computer to do any task which a human can do (unless it requires judgement, or skill computers do not have). "Generally – if you can do it as a human – we can automate as a computer", he said. "We can automate everything.

It is quite scary."

"And – by doing all this – instead of clicking manually – and do mundane tasks – you are having fun," he said. "If you love what you do – and do what you love you never work a day of your life. Your work will never feel like work, it is just playing games at a different level every day."

Doing projects like this one will also teach you more about the real world behind the data, such as, in this case, how geographical co-ordinates work. You can get some understanding about how websites are built.

Brunei Shell – moving documents to a standard structure

Brunei Shell embarked on a project to move technical documents onto a centralized structure, to ensure that they were accessible to the people who needed them

Brunei Shell recently embarked on a project to put technical documents into a centralized structure, so they would be available to the people who needed them.

Hjh Noor Raidah Hj Basir, subsurface and wells document records management (DRM) coordinator with Brunei Shell, explained at the Digital Energy Journal forum in Kuala Lumpur.

The subsurface and wells document records management team of Brunei Shell is part of the technical data management department but focusses on the "document side" of data. This includes making sure documentation is kept safe.

Before the project began in 2018, documents were being stored in multiple places around the company, including drives on personal computers, different SharePoint drives not accessible to the people who needed them. There were also multiple copies of documents in circulation around multiple departments, she said.

Sometimes a person would leave the company and no-one else would have access to their documents. Another problem was a lack of confidence among staff using documents created by someone who had left the company, because no-one could be held accountable for any problems with them. It was not easy to find out if the person was replaced by another person in the same role.

Documents were getting lost, or misfiled. A lot of time was being wasted searching for them. Sometimes people were having to 're-create' a file they could not find.

There were compliance problems, such as not being able to produce a project control assurance plan document when required.

People were asking for a means of proving that data was correct and could not find the document showing the calculation method and source. For example, asking, "how can you prove to me this TD (total depth) is correct if you don't have the document that proves it?"

The new system

The new system involves standardised structure of documents which would be saved for each project. The standardised structure includes a specific data storage location. Specific roles and responsibilities given to specific people to ensure that the documents were properly managed.

There is careful control over who had access to the documents, managed by the data management staff.

All Brunei Shell staff can see a list of all the different projects, but they can only access the documents in the folders if they are authorized. Many exploration sites are also locked.

When you are logged onto the main navigation page of the project site, you can see the names of all the other people who have access. You can search for information about different wells or fields. There is a form to fill in if someone needs to be removed or added, and guide to how to use it. The project owner approval is obtained before access is granted.

Hjh Noor Raidah Hj Basir, subsurface and wells document records management (DRM) coordinator with Brunei Shell

Documents

Once logged in, you will be able to view all the documents related to the specific project and direct link to the documents which has been declared as records. . The system is the same for both completed and ongoing projects. There is a "standard look and feel" for everyone.

Any final documents need to be 'approved' before they can be published as records within the site. The process is that upon approval, request for final documents publishing will automatically trigger an email to the publishing team for action.. The approval system and request are within the software, not handled by sending e-mails to people.

When a document is approved to be a permanent record, it is moved to the main records area and a URL link will be replaced in the original project site location.

The document management and the project leadership team can now easily monitor the progress of each project. There are controls in place to stop people copying documents or making them accessible to people who should not be able to see them.

There are now over 160 different project sites using the system. The company calculates that the amount of searching for documents has reduced by 50 per cent.

SharePoint problems

The new system has helped encourage some users to start using SharePoint, who were previously reluctant to move from storing documents on their own PCs.

The data management team holds a "clinic" every 2 weeks where people can discuss any problems with the system, she said. Most of the problems turn out to be SharePoint related.

The clinic is an opportunity to teach people "tips and tricks" on using SharePoint, such as how to upload multiple documents at once.

Geologix – using data to understand production

Geologix, a company specialising in analysing and working with E&P data, built a system to help an oil company better understand the reasons for production not achieving target on certain wells

Geologix, a company based in the UK, US and India specialising in working with E&P data, built a system for an international oil company to help it better understand why its production was not achieving its target on certain wells.

The client's head office was in Houston, with a drilling project in Africa, and a development team in Asia. So three different time zones, and three different cultures.

The client's objective was to maximise production – and if targeted production was not achieved, to be able to understand why, and to put measures in place to improve the situation.

"This was the requirement given to us. No design, no methodology, nothing. We started the product development from that stage," said Piyush Pandey, director of Geologix, speaking at the Digital Energy Journal forum in Kuala Lumpur.

Geologix developed a system which the well managers could use to monitor production from every well in real time.

There was a colour coded system where wells on track to achieve targeted production were marked green, wells which were not were marked red. So production managers could easily see if they were on track to reach targeted production.

This does not reveal anything about the underlying problems. But it is useful to get earlier warning that production in certain blocks is unlikely to reach target, Mr Pandey said. You can bring up data from specific wells, and see if the production rate is declining. You can compare it with wells nearby.

You can examine the well with declining production further, for example to see what the pressure is downhole in the well.

The causes for lower production can be unusual. With one well which was not achieving targeted production, Geologix discovered that a local company was given responsibility for collecting and observing the production data. This company would then communicate with the oil company's maintenance team when it observed a problem, so they could try to fix it. But there was quite a long time lag between a problem, such as low pressure observed at the bottom of the well, and a solution being implemented.

Geologix's system provided real time production data directly to the production department, so problems could be identified much faster, and solutions put in place to fix them much faster.

Value in data

Behind projects like this, a lot of thinking is needed into what data provides the most value to a company and how to present it in the most useful way, Mr Pandey said.

It is very common in oil companies that large amounts of data is generated about wells, and people keep records about problems, such as losing a drill bit, or not achieving targeted production. But they don't work out how to use the data to better understand the cause of a problem, and so prevent it happening again.

Piyush Pandey, director of Geologix, speaking at the Digital Energy Journal forum in Kuala Lumpur

For example, the data could show that a certain drill bit does not perform very well in a certain geology. If the company is drilling through that kind of geology again, and is aware of this problem, it may choose to stabilise the drill bit so it has less vibration, Mr Pandey said.

Any data has three dimensions to it. The point in space or the objects it derives from, the point in time it derives from, and what business value it could provide to the company, he said.

Lots of data is gathered and stored, but much less is analysed.

Geologix offers a range of cloud based business data aggregation and analytics services. It is developing new data visualization technology including with augmented reality, and digital transformation projects.

It is getting involved in decommissioning projects, which can turn out to have the same data management problems as production – because you need to get a good understanding of wells before plugging and abandonment, Mr Pandey said.

A data on demand system at PETRONAS

PETRONAS' technical data department has been developing a 'data on demand' system, so people can access data as they need it, rather than ask the technical data staff to find it for them

PETRONAS' technical data department has developed a 'data on demand' system for subsurface data, which will enable technical staff to access the data as they need it automatically, rather than ask the technical data staff to find it for them.

Chew Wei Liang, Manager, Data Operations Surface, Group Technical Data at PETRONAS explained how it works, speaking at the Digital Energy Journal forum in KL.

Before this was implemented, the Group Technical Data department were spending a lot of time responding to requests for data, which can often take several days to compile and deliver, he said.

Chew Wei Liang, Manager, Data Operations Surface, Group Technical Data at PETRONAS

Typical problems are that data is available but in the wrong format. Or the data is there, but they are unable to access it due to limited access rights. Or they can find lots of data, but it is impossible to know which version is the final version to use. Or the data is in the wrong format for the software they plan to work with. Or the data is available, but poor quality.

Now a system has been built to deliver data automatically 24/7.

Right data is available to people as they need it, directly out of corporate databases – so no manual hunting for data, or data checking, is required. People can obtain all the required data in a single platform. You will have assurance that the data has been quality controlled.

You don't need to know where the data is actually sitting, you just need to know what data you want.

It means that the role of the data management staff changes to being a custodian of the data, rather than loading it and delivering it to people, he said.

The data management team can help get files organised at the start of a big project, but after

that, access to the data by the users is "self-service", he said.

If the data is not available from the central storage, it means either the data was never acquired, or someone forgot to load it to the database. This provide full transparency to the users.

"We can allow the user to download the data in the format they want" he said. Although there are some compatibility issues between some database systems and some software systems that can be resolved by adopting industrial standard such as Energistics.

The data platform can put all the data in technical context, for example tying together data about a platform, with a field, with projects. So, you can click on assets and get a list of data relating to the particular context.

The quality control process can involve asking subject matter experts to verify the data is good quality by creating biz rules for each data type. The data management staff can monitor if it follows the business rules.

You can see a colour code indicating the quality of the data, or whether or not it has been quality controlled yet. A grey code means the data is not available, yellow is that it is available but not quality controlled (QCd), green means QCd. If data is updated, green can turn to yellow.

Data types

The data is structured around the needs of different departments, such as exploration and production.

There are over 280 data types across exploration, development, production and downstream – of which 224 is in upstream and 56 downstream. The downstream data is much more standardised than upstream, "basically engineering data, sensor data, maintenance and inspection reports, HSE," he said.

Managing unstructured data, such as from spreadsheets and data in pdf or slides, can be a challenge. "We call it 'homeless data' – we try to create a home for it."

Managing downloads

Regularly used data can be stored in a caching layer, rather than being drawn out of the database every time. For example, production data, which is commonly requested, can be updated daily to the caching area.

Some well data needs to be kept secret, such as certain exploration wells (known as 'tight wells'). The restrictions can be managed within the system.

Altogether, 150 GB of data is downloaded every day. The data management staff can monitor who is downloading what, and check that people are downloading data relevant to projects that they are working on.

It was not the intent that users would "stream" data to their laptop as they need it, more that they would download data they need and store it locally.

"It is possible to monitor all the requests and where they come from. If you find any discrepancy, we block the users or service account."

One project team was downloading 1GB of data a day. This was blocked because it sounded suspicious, he said. "There is no need to download so much every day."

It is important to look at the data types not just the file sizes, because someone might be downloading large seismic files justifiably.

Machine to machine

A future step is called 'data gateway', which will involve much more machine to machine integration via APIs.

It will mean that someone will be able to access the data within their software, through web services, and it is made available, without any human intervention.

PETRONAS has created APIs for multiple types of applications, so you can get direct access to multiple databases.

To date, the data platform has enabled PETRONAS to have a cost saving of approximately RM200m ($48m) and reduce the time to data from days to minutes, he said.

Sarawak Shell – condensing subsurface data onto a grid

Sarawak Shell has a project to integrate subsurface data from multiple surveys and interpretations onto a single grid, covering North Borneo, and only showing the best data. Senior technical data management consultant Teck Hing Wong explained

Sarawak Shell has embarked on a project to integrate together data from multiple seismic surveys and interpretations into a single grid covering North Borneo, showing only the best data.

The aim is to put everything together in a single grid, which would provide the exploration team with "the final trusted exploration data set at their fingertips, with all data synchronized," said Teck Hing Wong, senior technical data management consultant with Sarawak Shell.

It should help reduce the amount of time users spend searching for data, ensuring that all data is in a central corporate database, rather than on people's hard drives or within corporate silos.

The integration project also involves stitching together data from multiple seismic surveys. Stitching together seismic data is a complex task. Seismic is not always shot at the same grid pattern or angle. There can also be inconsistencies in the seismic sampling rates and spacing,

Teck Hing Wong, senior technical data management consultant with Sarawak Shell

so missing data points, he said.

The project will also bring together data integration work, although it can be made at a wide range of scales, from large scale "play" interpretations to project interpretations.

Data quality

In order to produce this, you need controls over the quality of data which is entered into the system, so data is not pushed out to anyone unless it passes a certain standard.

Understanding data quality requires a level of technical expertise which data management or workflow support staff do not usually have. So, the initial data owners need take some responsibility for data quality, he said.

You need to set the data quality bar at the right level. If it is too high, you reject lots of data, in which case you don't have the data you need in the database. If the bar is too low, you accept too much data, and ending up with too many versions, he said.

The quality of the data, and whether the data is in the right place, can be far more important issues than the digital technology you use, he said.

The project team decided to constrain the number of data types it would look at to 10. Important data types included well top data, log curves, check shot data, seismic data, seismic velocity models, and horizon data.

The seismic velocity models themselves needed to be integrated together. This is very difficult because they are often put together for specific tasks, and put together in different ways, with different audit trails.

Business driven

The data management and workflow support staff wanted to start a project by asking the business owners and system users what they wanted.

It is a temptation for data management staff to "try to fix everything" without any focus, he said. Data managers can have a mindset of "my job is to make sure I manage the data, get data complete in the database". But this makes a project too big.

Instead, it is better to ask users for very specific information about what they need fixed. Which specific area they were interested in, which wells they want data managed for, and when they need it.

Technology

In terms of technology, the company has been looking to build up basic digitalisation capability, encouraging people to do basic coding and scripting, and built the digitalisation culture in the company.

It is aiming to recruit graduates who have this sort of knowledge – a "hybrid of geoscience knowledge and computing."

It is looking at improving visualization, with tools like Spotfire and PowerBI. "We want the business to have better visualization," he said.

There is no single database which can store all the data – you need a different database for different data types. It makes it hard to build a platform to bring it all together, so it is easier for people to find.

The company has software tools which can highlight anomalies in a large area of seismic, perhaps showing them up in different colours. It can bring out different features such as channels.

There is machine learning software which can take a full stack seismic cube, filter out the noise automatically, and highlight likely faults using pattern recognition. This could take an interpreter two weeks to do. "The quality of a tool is not as great as how a human would pick it, but for a regional prospect this is great," he said.

Energistics – using RESQML to move subsurface data

The RESQML data exchange standard makes it possible to move subsurface data from one software application to another, with no loss of data fidelity. This makes it possible for geoscientists to adopt much more complex workflows with multiple software applications made by different companies, says Ross Philo of Energistics

The RESQML data standard makes it possible to move subsurface data models from one application to another, with no loss of data quality, and maintaining records of everything which has happened to the data along the way.

This makes it much easier for geoscientists to adopt more complex workflows, making use of a combination of software applications, each of which might be focused on one specific task, rather than having to do everything on one universal software application, said Ross Philo, CEO of oil and gas data standards organisation Energistics.

He was speaking at the Digital Energy Journal KL forum in October, "How to Digitalise Exploration and operations".

In the past, the development of earth models was a highly linear process, first creating a structural framework of the subsurface in a cellular grid, then adding in reservoir properties. Geoscientists could do all of this on a single software package, designed to work in this linear way.

Ross Philo, CEO of oil and gas data standards organisation Energistics

But now, people want to do more and more steps, and not necessarily linear steps. For example, they might want to put in pre-stack seismic interpretation, add geomechanics interpretations (studies of rock stresses and how that affects seismic properties), or do some analysis on a subset of the data.

Geoscientists may wish to add in alternate grids or bring in other types of data analytics or machine learning, or chrono-stratigraphy (adding geological time to the identified rock layers).

The reservoir production itself is becoming far more complex, as companies look at methods like enhanced oil recovery (EOR) and inject water and gas.

The applications have also become far more comprehensive. Companies might want to move the entire model into another application to add "additional depths to the analysis" and then move it back.

There are a number of smaller companies developing solutions for specific tasks, which companies would like to incorporate into their workflows, but can't because the challenge of moving data around gets too great.

Normally, the only way to do this is to export very large files from one software application to another. You would need a way to link each software application with each other application. It gets very difficult managing the relationships between all of the files and maintain data integrity.

"You end up with a cat's cradle of complexity that becomes incredibly hard to develop, manage and maintain," he said.

There is a limit to how complex you can allow a process to be – with every additional applica-

tion adding more complexity, he said.

Sometimes work is re-done because someone is not confident it has been done properly by someone else before.

"No single vendor can cover the complexity of all the different workflows companies want to use today. You're looking for best of breed applications that you can mix and match in order to provide the sort of flexibility that is required," he said.

"You need to have a way of plug-and-play for combinations of solutions, so you can move data from one application to another as you need to."

There are also more and more data types. For example, companies are making more use of distributed acoustic sensing (DAS) systems, with fibre optic cables in wells, which can generate up to 10 terabytes in just a day.

Keeping track of data objects

The technical challenge with RESQML is to be able to move earth models from one application to another, where it can be unpacked by the receiving application without losing any fidelity.

Subsurface data can come in a range of different co-ordinate reference systems, means of measuring depth, and units. If there is any mismatch, the result is a big mess.

The subsurface data model must keep track of every object, in the right geographical location, including horizons (where you think the rock layers change). Individual objects need to have unique ID numbers.

You also need to keep track of what has happened to data in the past. Mr Philo uses the analogy of keeping records of a piece of art, how you prove who made it, and what has happened to it since then.

It gets very difficult to keep track of different components and their relationships. Similar to how it can be hard to keep track of data relationships if you let someone else work on your spreadsheet for a while.

The data files can get very large, which also makes them trickier to move.

Demonstration

A demonstration was made in October 2018 SEG event in Anaheim, with data for a field jointly operated by BP and Shell, moving data between 6 software packages, including from Roxar, Paradigm (both now part of Emerson), Petrel, and a special fracture porosity software (OpenFlow) from IFP, reservoir simulations from CMG, and a final visualization in Dynamic Graphics.

The workflow was designed so that the reservoir model could be enriched by specialist software tools.

In the project, a subset of data was processed in OpenFlow and then passed to Petrel. It means that people don't need to work on the whole model, they can enhance a subset of the information and re-integrate it to the original model.

The demonstration also showed how earth model data could be moved with complete fidelity between applications running in different cloud environments as part of the overall workflow. The whole demonstration took 45 minutes and was demonstrated live at an exhibition stand. "They did this step by step and were able to update a set of the information and move it across to another cloud and bring it back," he said.

You can get the output with just one file, not multiple outputs from different applications.

It makes it possible to have more people collaborating on the work, and increase the type of analysis you can do, and reduce the risk on exploration decisions, he said.

Also, as the data was moved, it was possible to keep track of every element in the earth model, including the history and lineage of data.

Data archiving

The standards are designed for data transfer, moving data from one application to another, but can also be used for data archiving,

By putting data in open data standards, an oil company has a higher likelihood of being able to use the data decades into the future, than if the data is stored in a proprietary format.

Oil companies often find they cannot work with data which was created in software one version behind the version they are currently using, he said. Schlumberger has said it will only support 2 versions back. "You've spent billions of dollars on this data- you want to get to it again."

Provide data to regulators

RESQML can be used to provide data to regulators. As an example, the UK's Oil and Gas Authority has said it does not want its own copies of all the earth model data, but wants to be able to ask an operator for data about the history of a field at any time.

This might include data about the original decision to develop the field, and data from a number of different companies who owned it along the way.

If the data is not stored in a standard format, it probably means that it can only be used with the software application it was originally created in.

"I would love to see regulators requiring data in the standard format," he said. "They are already asking for PRODML for production reporting, I think we'll see them requiring RESQML for earth models."

OSDU

The Open Subsurface Data Universe (OSDU) project is "really exciting – it seems that every conference we go to, the conversation is all about OSDU," he said.

OSDU is designed as a standard data platform for subsurface and wells data, including seismic data in a 2020 release. It aims to keep data decoupled from the application. There is a series of APIs which will allow software tools to connect to the underlying data in a standard manner. Energistics was one of the first non-operators to join in November 2018, since it made sense that a standard data platform would include support for Energistics data exchange standards.

It has many of the largest operators signed up to it, and represents a highly-collaborative effort, involving a large number of companies within the OSDU community, to deliver the solution. "I think it has tremendous momentum, and shows what can happen when a group of operators get serious. The proof will be in how it then expands to attract other operators and service companies, as well as being able to cover other data types." As of Dec 2019, there are 26 operators involved, and a total of almost 120 members.

It is not the first time that attempts have been made to develop an industry standard integrated data model – other projects include POSC and Openspirit.

But this one is different because it makes use of the cloud, he said. "The previous platforms proposed would have required an operator or service company to implement that platform within their own environment. With OSDU – the intention is that each of the major cloud providers will offer it as a service which an oil company can then contract to."

About Energistics

Energistics sees itself more of a 'custodian' of standards rather than an organisation which develops standards. "The standards are defined by subject matter experts within the different companies that are members of Energistics and are freely available to the industry," he said.

There are 115 members, including oil companies, service companies and software companies who collaborate to define the standards and to support Energistics' activities on behalf of the industry.

The three major standards are WITSML, covering drilling and well construction, PRODML, covering production (production volumes and the production string), and RESQML, covering earth models.

Clustering considerations for machine learning

Most datasets in oil and gas are multi-dimensional having many variables that make it difficult for us to analyse and find meaningful patterns. Therefore, the reduction of dimensionality is a fundamental part of a machine learning workflow and cluster analysis is one of the key tools used for this. Different dimensionality reduction and data clustering techniques are available.

Philip Lesslar, a data solutions consultant with Precision DM, formerly with Shell and PETRONAS, explained some of the techniques, speaking at the Digital Energy Journal forum in KL in October, "How to Digitalise Exploration and Wells".

The preparation of data for any kind of analytics or machine learning starts first with reducing the dimensionality of the data and that can be done using a number of different multivariate statistical techniques.

He started first by describing a higher level group of dimensionality reduction statistical techniques.

a) Cluster analysis is a technique that aims to find "natural" groups in multivariate data sets,

b) Principal Components Analysis looks at reducing dimensionality by finding a smaller set of variables that is still representative,

"Principal component analysis" can be used to reduce the dimensionality, identifying which variables make the biggest impact on others and which seem unrelated to others. So, for example you can reduce the number of dimensions you are working with from 100, which is very hard to make sense of, to 10.

You want to preserve the "information" – the useful signal – while reducing the volume of data. An analogy of this is when we compress files to zip format. There are different types of zip – "lossless", which preserves all the information, and "lossy" which tries to only preserve critical information.

"So, we can think of this as like something we use already," he said.

c) Factor analysis is useful for datasets where a large number of observed variables are thought to reflect a smaller number of unobserved variables,

"Factor analysis", similar to "principal component analysis", can be used where you believe a small number of variables – perhaps unmeasured directly - drive a large number of other variables.

d) Multi-dimensional scaling is a technique that helps visualise similarity of samples by transforming onto a 2D plane, and

e) Linear and multiple regression are techniques where one or more independent variables are used to predict the value of a dependent variable.

He said that the aim of this talk is to just focus on cluster analysis and its significance on the machine learning workflow. He then described the general features of cluster analysis and some of the key types available:

Cluster analysis is a methodology for classification of objects with many data points. People have been doing classifications long before we had computers. For instance furniture is a class, and chairs, tables are subclasses. We recognise a chair when we see one even though it may look different from what we are used to. Our brains have learned the features that a chair possesses. Another example is Charles Darwin classifying organisms in the 1800s. Classification is part of how people learn about how something works, and part of how machines learn, Mr Lesslar said.

When working with data that have just two or three variables, it is easy to plot the data and visualise the groupings. Classification gets much harder when we are working with data that have many variables (multivariate). This type of data is harder to visualise e.g. seismic data, or training a computer to analyse an image. We often have to look at data in many different ways to get insights from it. All data sets contain "things you can easily see, and information you don't see until you transform some of the data," he said.

"Data analysis is not for the faint hearted," he said. "If you are a dabbler, dabble yourself out of it."

Most machine learning projects start with some form of cluster analysis – the first step is to create meaningful groups out of a collection of objects (classification), and the second step is to build a model about how the groups behave, based on extracting features out of each group. Then in future the model can be used to identify which group a new object belongs in.

Definitions of machine learning workflows talk about the "training data set" and the "testing data set", which you use to build and test your models. It is important not to use the same data set for both.

Machine learning gets much harder the more variables you have, because you don't know which variables are most important. And there are many variables in exploration and production, he said.

Philip Lesslar, a data solutions consultant with Precision DM

Cluster analysis techniques

One popular visualisation-based clustering methodology for data sets in machine learning is K-means, which allocates every data point to the nearest cluster based on its affinity to the mean of that cluster.

Mean shift clustering is an iterative method where the computer looks for the "highest mean density" of point groups.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is similar to mean shift clustering but is better at spotting outliers.

"If we take the complexity out of it, relate it to what we already know, then it becomes less of a mystery," he said.

Another common method is Expectation–Maximization (EM) Clustering using Gaussian Mixture Models (GMM). This method looks for clusters where data follows a bell curve, or simple normal distribution. This detects elliptical clusters, not just circular clusters around a mean (centre).

"Agglomerative hierarchical clustering," is progressive pairwise clustering or finding the two points with the strongest affinity, such as someone entering a bar and talking to the person they have the strongest connection with. They then form a two-point cluster. A third person, a friend, joins them and now they form a three-point cluster. "You are carrying out cluster analysis in a bar," he said.

The results of cluster analysis vary if you look at the data according to different dimensions. For example, individuals could be sorted according to their age, or home town. The "hierarchical clustering" method finds the cluster dimension which makes the most sense.

In summary, there are two critical elements to your cluster analysis – one is the similarity measure used to calculate the "closeness" of points in n-dimensional space, and the other is the clustering algorithm that calculates the progressive clusters.

Proximity measures - Relations between data points

He outlined a number of common proximity measures that are used to calculate "similarity" of data points based on various attributes that these points possess. These attributes can be either quantitative (measurable) or qualitative (nominal). "Many of these measures have been developed years ago" he said, "an example being the Jaccard coefficient of similarity which was developed in 1908".

Although these indices look mathematically complex, they can be more easily visualised and understood using Venn diagrams and set theory.

A particular measure, the Euclidean Distance coefficient, looks complex in its n-dimensional form "..but if we reduce the dimension to 2, then we have Pythagoras's Theorem for right angled triangles which we have all studied in secondary school."

Some examples from exploration data

In order to illustrate the use of cluster analysis on real data, Mr Lesslar showed examples using prospect appraisal data, well logs and micropaleontology (foraminiferal sample data).

Prospect volumetrics typically include attributes such as POS (probability of success), MSV (mean success volume), HSV (high success volume), REC (recoverable), STOIIP (stock tank oil initially in place etc. By using these as input, one can expect that similar prospects will cluster together. He showed the resulting dendrograms in which some clusters were clearly seen and were consistently seen even using different clustering algorithms. Without any prior knowledge about these affinities and clusters, results may sometimes prove surprising and trigger further ideas.

In the example with logs, Mr Lesslar simply took the digital point values of several log types in a well section and ran them through a few clustering algorithms. Similar to the previous example, a number of clusters could be clearly seen. The key point in the exercise was to show that patterns were there but would depend on further work to ascertain whether these patterns were significant or meaningful.

The last example made use of foraminiferal assemblage data in well samples to show that these data lends itself well to cluster analysis. Clear groups could be seen and it is well known in micropaleontology that foraminiferal assemblages are environmentally sensitive and such clusters can be used to identify groups of environmentally similar assemblages.

When doing cluster analysis work, you might try various different techniques on your data set and see what happens. You might see some clusters which look particularly interesting, and look at them more closely.

"We make no assumptions about data, let's just explore," he said.

The clusters may not be immediately obvious, but reveal themselves with cluster analysis methods.

If you see certain patterns appear, you can bring in domain expertise to try to understand if there might be any sensible meaning behind them. Perhaps an expert might suggest looking for clusters around a certain depth, because it is in a different formation.

"Some parts we may never understand, they may be so complex. But at least if we know 80 – 90 per cent of moving parts, we have a better chance of getting meaningful results from this technology," he said.

Oil explorers have used clustering techniques with micro fossil data to try to identify patterns in the groups of fossils found in different parts of a well, and how that relates to other factors, such as presence of hydrocarbons.

By tracking the patters in the micro fossils, you can try to map the migration path of oil back to the source rock. Or track different source rock types, source rock maturity, and the temperature and pressure of the source rock and sedimentation rate. And ultimately you can use it to find new source rock and identify if it might have generated oil.

Brunei Shell - monitoring compliance of wells data

Brunei Shell has developed a Spotfire based tool to monitor compliance of its wells data with the company standard. Nurhamizam Yussop, wells and production technology data technician with Brunei Shell, explained how it works

Brunei Shell has developed a Spotfire based tool to monitor whether its wells data is in compliance with the company standard. Nurhamizam Yussop, wells and production technology data technician, explained how it works.

Nurhamizam Yussop has been with Brunei Shell for 11 years, initially as a mechanical engineer, assisting with production technology, then moved to the technical data management team for wells and production technology in 2015. He has been involved in a number of other Spotfire projects in subsurface, logistics and change management.

Nurhamizam Yussop did a project to see if it was possible to automate the quality control running Spotfire directly on some of the reports, such as the casing report. If you know what tables and columns you are looking for, you can write script to run in Spotfire, to tell you if that data is available.

You can quickly pinpoint errors. There is no need to open up each report separately.

Web engineers can also directly access the dashboard through Spotfire's web client.

As an example, you can see all the rigs, which wells are currently active from each rig, and the reports which have been provided so far. If everything is available, the well is shown as green on the dashboard.

It is possible to look deeper into the causes of problems, such as a problem reporting casing in the Engineering Data Management system.

There are logic rules, such as that the outer diameter of casing or tubing should be larger than the inner diameter. If the inner diameter is larger, "it will say no".

The "proof of concept" of the project was completed in summer 2019, on the Spotfire development server, and will now be moved to the production server.

Nurhamizam Yussop, wells and production technology data technician with Brunei Shell

Basic process

The company operates a number of rigs, continuously drilling wells. Well site engineers are asked to leave copies of their wells data and documents in the company's standard wells database, EDM OpenWells software from Halliburton

There is an "Operating Reporting Standard" in the company, to state what data is needed for each well.

There is also a checklist of everything the wells team need to send.

After data has been deposited in the Engineering Data Model (EDM) the data flows to a "staging area". The data team need to check the data for anything which needs correcting.

The company's technical data specialists do checks of the data for quality and completeness. Checks include whether there is data about the casing and primary cementing, the base of cement, the leak off tests, the well properties. Altogether there are 100 fields to be checked.

If anything is missing, for example a missing attachment, comments can be entered, and it is sent to the rig team.

Well site drilling engineers normally work 2 weeks on, 2 weeks off. If requests for amendments are sent to crew A, sometimes you have to wait another two weeks until Crew A resume work on the rigs, he said.

Once data about casing, cementing, completion and other aspects is in the system, it can generate a well status diagram. This is sent to the completion engineer to check, and then sent to the well owner for final approval.

Finally, the well status diagram can be handed to the well activity owner and used during the lifecycle of the well. The wells data is also used by a number of other applications.

It is important that the data is accurate. If the company ended up with incorrect well data, it might plan maintenance wrongly, he said.

digital energy journal

How to make data science useful in oil and gas

Poh Hean Yap, senior manager with Accenture, explained some of the best ways to make data science useful in oil and gas – including by combining data science expertise with domain expertise, and focussing the work into short 'sprints' to develop a minimum viable product – with a case study of how data science saved millions of dollars

Poh Hean Yap, senior manager with Accenture, explained some of the best ways to make data science useful in upstream oil and gas, including by combining data science expertise with domain expertise, and working in focussed 'sprints', based around a 'minimum viable product', so the client can quickly get a return on investment.

She presented a case study of how Accenture saved an oil and gas client millions of dollars, with a tool to better manage its injection of chemicals to manage emulsion in oil, and recommendations to change the sampling points on an offshore platform.

Ms Yap has been combining data science with domain expertise since doing a Phd in data mining for the textile sector. Her Phd project was to identify factors which impacted the fibre quality of wool in Australia.

Wool which does not meet the fibre specification in Australia cannot be sold for use to make clothing, so "your whole batch needs to be torn up and made into carpet," she said.

To understand fibre quality, she had to develop an in-depth understanding of how the industry worked, and the factors impacting fibre quality. This was then used together with data science to make recommendations.

Ms Yap has also spent 6 years working in a refinery, just focussing on using data for optimisation. Working for such a period in a single domain is a good way to build up domain expertise, she said. Now, she focusses only on data science in the resources domain, including oil and gas, chemicals and mining.

Ms Yap's case study, outlined below, was a system Accenture built to enable an oil and gas company to better understand how much chemical it needed to inject in an oil processing facility, in order to prevent emulsion forming in the oil, and oil subsequently getting rejected by the client, a refinery. In order to do this, you need some knowledge at least of what an emulsion is, she said.

Some data scientists promise just to be able to make predictions out of data, with no domain expertise. "I tell then, go away. If they don't try to understand your domain before they start telling you what can be done, it is really useless."

A minimum viable product

It is easy for data science projects to get very long, lose focus and become very expensive. To avoid this, Ms Yap advocates that projects should be constrained to a finite length, such as 14 weeks, and aim to deliver a useful output within this time, known as a 'minimum viable product' (MVP) which is useful enough to provide a return on the investment

In the industrial data science world, the MVP might be a dashboard tool for calculating or monitoring something specific, plus some useful recommendations.

Ms Yap explains the concept of MVP with the analogy of someone developing a device for personal transportation where none had previously existed. Your MVP in the first stage might just be a board with wheels, like a skateboard. You might add a handlebar to help balancing in the second stage. You would not try to build a car from scratch.

Poh Hean Yap, senior manager with Accenture

Another analogy is how Apple developed the iPod before it developed the iPhone, she said.

A focus on MVP also means you are focussing on collecting the minimum amount of data you need to make something which works. This is important, because otherwise you can spend enormous amounts of time finding and cleaning up data.

"Let the data scientist tell you if there is enough data to work with or not," she said.

Starting a project

Projects often begin with a vague request from the client, such as "show me how digital can help us", "show me how I can resolve this specific problem with digital technology", or "can you build an AI which can scan through documents to tell me, how much should I be spending on this well."

In the example below, the initial request from the client was, "we suspect over-use of chemicals [for injection], we are spending too much money, can you please look at it."

So one of the most important parts of the work is aiming to distil this initial request into something specific which may be possible to deliver.

Accenture takes the entire first two weeks of the 14 week period, doing a workshop with the client to try to identify in depth what the client is trying to achieve, and where to start.

Participants can be split up into groups, each discussing how they think digital technology can help them, or what they want to achieve. "You have to define your North Star," she says. You need more than "I just want to go digital".

Participants are asked to specify how the company would get business value from the investment, such as from achieving a 10 per cent reduction in spend on wells, or reduce the cost of optimizing wells by 5-10 per cent. They can't just stipulate software as the desired output, such as "I want a dashboard".

Accenture brings in a number of use cases from similar projects it has done.

This process may generate a number of ideas, which need to be further screened and refined for their business value.

You need to consider how much the project will disrupt the organization and how much value it will achieve. If you need to change the entire company to do one thing, it will need to show a lot of value to be worthwhile. Conversely you may have projects which are high value but low disruption, so they are relatively easy to do. "That is your low hanging fruit," she said.

You also need to consider that the people actually attending the workshop may not be representative of the whole company – there are people who did not attend the workshop who will not gain much value from the product, but may also need to participate for it to be successful.

Use on other equipment

Companies are often tempted by the idea that they can develop a tool which can then be used in multiple different ways around the company. "They've paid for it, they want more value from it."

But this rarely works. "All the data is very sensitive to wherever you are trying to explore," she said.

You find that parameters work together in different ways on different equipment, or something else matters.

Even if the model is being used on very similar equipment, it will require re-training and customisation to work on the second set of equipment.

Work process

Once you have defined what you want to achieve, with a rough idea of the 'minimum viable project', the work can begin.

The first step is to define the team structure and governance system.

Typically, there will be a "business results manager" whose role is to keep the focus on making a MVP and support the work of getting there. There will be business "domain experts" who understand the industrial process the system relates to. You can have data scientists and data engineers to work with the data.

You may have business analysts which sit between the domain experts and data specialists. And you may have visualization experts who drive how the data should be presented and user interface should work.

You can have 'scrum masters who can stream the work if it is in multiple streams.

You have to define what technology you will use, for example Spotfire for visualization.

Once the direction has been defined, the next 11 weeks of work can be very focussed and fast paced, experimenting with different ways of working with the data and discussing it with clients. The work with clients is collaborative. "We want to bring you on a journey with us", she said.

Much of the 11 weeks work includes pre-processing and cleaning of data.

The "data discovery" work may give you some results, for example a means of predicting what the output will be with certain input data. Then you can build this into a model, put it into a software tool, which the customer can use to predict what will happen based on the same given input data in future.

There is a difference between "predictive" and "prescriptive" – prescriptive tells you what is actually happening or going to happen, predictive has elements of probability in it.

The 11 weeks allows time for "deep dive" into the problem. Perhaps you will find some people in a company see a problem, but not everyone agrees.

Chemical injection example

Ms. Yap presented one example, of an oil and gas operator which felt it was spending too much on chemicals which were being injected into the oil stream to stop it forming an emulsion.

The oil had been rejected by a refinery downstream because it was too high in emulsion, leading to the injection rate being increased.

The client started by just saying, "We are spending too much money, we suspect over-use of chemicals. Can you please look at it."

At the process plant, the work process was to take a sample of the crude to see if it was forming an emulsion, and if not, to add more chemical. The oil needed to be heated to a certain temperature for the chemical to work.

Building a model

In data science terms, this project could be defined as a request to make a model of how much chemicals are actually required to be injected in order to ensure the output oil meets the specification of the downstream refinery – and then assess how closely this matches what is actually being injected.

A first task was to discover how much chemical was being injected. Accenture's team discovered that the client had no reliable records.

The chemical injection rate could be recorded by operations staff in handwritten notes, which were then typed into a spreadsheet and sent periodically to a production chemist. Sometimes there were gaps in the data, which could indicate that someone had forgotten to record the amount of injection, or nothing was injected.

Sometimes people offshore forget to inject one day, then realise, and inject twice as much the next day, but do not record this omission in their records.

Accenture's team also found out that there is very little communication between the various silos of the company. There was a downstream group which buys crude from the upstream group. The downstream group says the oil does not meet specification, and causes the refinery to shut down. The upstream group claimed it did meet the specification. There was not much communication between the production chemistry specialists and the separator maintenance teams.

The first discovery was that the oil from the upstream group actually was not meeting the specification, but still being shipped.

A second discovery was that the impact of chemicals depended on the specification of the input crude, and the operating conditions. So to build a model of how much chemicals were required, it would be necessarily to understand how the system is working.

How the system is working

So Accenture's team needed to understand how the oil was actually flowing around the processing plant.

A simple process flow diagram was drawn, showing that incoming crude goes into one of

four heat exchangers, then goes to separators, with some circulation so it would then heat up more incoming crude.

The team found that the client's existing piping and instrumentation diagram was inaccurate, with sampling points positioned in the wrong place.

The chemical was only injected into two of the four heat exchange streams, so half of it was not being treated. But the sampling points were in other two flow streams – so the company was not injecting and sampling in the same stream.

Build a data model

This leads to work to construct a data model about how the different functions inter-relate.

You need to separate causation and correlation in doing this. For example, you may observe that the emulsion level in export crude is low when a certain parameter is high. But this does not tell you that you have identified the parameter driving the emulsion level, she said.

At the end of the work, a data model could be built containing the predictions, constructed within a dashboard software tool for the client.

Given the inputs, the model can make predictions of what the temperature or emulsion level will be at a certain point, or where it will go in the next hour. You can ask the model, if you add a certain amount more chemical, how that will change the output.

The software tool was given to the client together with recommendations to move the injection point or use a different sampling point.

The system was then put through a field trial, where the client can be asked to close off certain trains, add certain chemical, and then compare the actual results with the modelling results.

LR - establishing 'safety tech' as a technology category

Lloyd's Register has a mission to establish "SafetyTech" as a technology category similar to "FinTech" – hopefully encouraging large amounts of investment and innovation into the sector. It showcased some interesting companies in an event in London in October

Risk and technical services organisation Lloyd's Register has a mission to set up "SafetyTech" as a class of technology, similar to "FinTech", hopefully driving innovation and investment in technology to improve safety.

It held a "Safety Accelerator Showcase" event in London in October 2019, as an opportunity for delegates to get immersed "in digital safety and risk innovations and entrepreneurship."

LR's "Safety Accelerator Program," aims to match corporates which have specific safety technology needs with technology companies. Corporates can set challenges, or specific problems they would like solved. Technology companies are invited to submit proposals for how they will solve them. The Lloyd's Register Foundation gives a grant to pay for the pilot project.

The Safety Accelerator is run in partnership with Plug and Play which claims to be "the largest global innovation platform and most active venture capital fund in Silicon Valley". It is then able to invest in the companies.

"We end up with an ecosystem and a community. tech company and corporates," said Steve Price, LR Accelerator Programme Lead. "Our job is to introduce people."

One idea for safety technology could be an analytics tool which can help company safety departments get more insights from incident data reports. There can be too many for a safety manager to read them all, so perhaps they could just see the trends.

Another idea for safety technology could be maintenance software which advises on what maintenance does not need to be done – rather than just tell us what maintenance is due.

There are a number of barriers to running pilot projects with safety technology, Mr Price said. It is very difficult to calculate the return on investment in safety technology. Also running pilots can be quite expensive, particularly when the business case is not clear. Projects can easily get de-prioritised behind something which feels more immediately important.

Sensing Feeling

Sensing Feeling (www.sensingfeeling.io) of London makes tools to "sense" human emotion, using video analytics. It is working together with shipping company Scorpio Group. The idea is that it might be possible to analyse how seafarers are feeling.

The company did not initially envisage that shipping would be a market for the technology, said Jag Minhas, CEO and founder. It was anticipating more interest from sectors like retail, events and hospitality, where companies might have an interest in sensing how people feel.

For Scorpio Tankers, the company is aiming to create a "risk index", mapping seafarer emotional state to risk of giving them control over a vessel, collecting data from real and simulated voyages.

There are questions of how relevant a model developed for one person would be to another person, since every individual is different. The system does not collect any personally identifiable data, for confidentiality reasons. But it may be able to adapt the models to take into account traits of people with certain gender or culture, the spokesperson said.

Ohalo

Ohalo (www.ohalo.co), based in London and Sunnyvale, California, is developing a tool to automatically anonymise health and safety data, so it could be useful for further analytics studies without violating anyone's confidentiality.

The UK Health and Safety Executive has 1.5m accident reports, and would like to share this data with data scientists, to see if they can spot trends in it, said Kyle DuPont, CEO and co-founder.

To anonymise the data manually was estimated to take 70 man years.

The system can automatically identify data such as name, company and address, using custom data classification models. It can then automatically remove (redact) it.

There are options about how much redaction is most appropriate, for example it may be useful to the data analytics to have some geographical context, and a decision can be made that this is not too much of an issue with confidentiality.

The auto redaction has been compared to manual redaction and is proven to be more accurate in some places, for example the computer may spot first names in the text more accurately than a person does.

M Squared Lasers

M Squared Lasers of Glasgow, UK, is developing laser technology which can analyse food condition when it is in a sealed bottle. Light interacts with different substances differently,

The LR SafetyTech event in London

getting reflected and absorbed in different ways, so a model can be built of how food in different conditions will interact with the laser in different ways. The laser can be "tens of metres" away from the bottle.

It helps that the company "makes one of the world's purest lights", said Robin Head, innovation laser engineer.

Senseye

Senseye of Austin, Texas (https://senseye.co/) is analysing the way that iris muscles (which control the size of our pupils) are connected to our minds. Scientists have long known that our pupils become larger (dilate) for many different reasons, which is interesting but not so useful for analysing the brain state. But if you look at the muscles which pull the iris (which controls the size and diameter of the pupil), rather than the pupil itself, you can get much more insight, says David Zakariaie, founder of Senseye.

The company aims to measure individual muscle fibres.

A trial is being run with Pacific International Lines, a container shipping company based in Singapore, where crew are asked to take an "iris test" before starting work. In theory, if the system detects a reduced psychological state (for example from fatigue), a seafarer could be assigned to a less dangerous job.

The system was installed in a trial running from June to October 2019.

It developed a training model, giving people slowly increasing amounts of alcohol to drink in a laboratory, verifying blood in their system using a breathalyser, and seeing how that affected their iris muscles. But the ambition is that the system will be much more than an expensive breathalyser, it will be able to assess people's psychological state.

The head of safety at Pacific International Lines is a former investigator at Singapore's Transport Safety Investigation Bureau, Mr Zakariaie said. In this role, he observed that a number of accidents occurred because of the individual was not in a state of mental fitness. This mental fitness state was then linked to a specific cause, such as bad news from home.

There are three stress factors where most people's irises respond in the same way – alcohol, drugs and fatigue. But for more complex stress factors, individuals respond differently – so you would really need to develop models about individual people to identify their mental state, he said.

In this, the computer's capability is similar to a human's – we could probably detect a stranger's alcohol intake, drugs intake, or fatigue by looking into their eyes if it was extreme enough, but we would need to know them personally to recognise their mood from looking into their eyes.

Speakers at the LR SafetyTech event in London

Another challenge is that the computer modelling uses a large amount of computer power – which can be hard to make available onboard a ship, Mr Zakariaie said.

Senseye has also worked with the military, testing the iris muscles of troops after and after deployment.

Allergy Amulet

Allergy Amulet, based in Wisconsin, is developing a testing system which people with allergies can use to test out food where they don't have complete information about the source, such as in restaurants.

It has developed a small electronic device, which analyses a sample of food.

Ms Barnes's motivation for developing the company comes from having a food allergy herself and having a "near death experience" accidentally consuming food containing an allergen. She is a lawyer and her co-founder is a chemistry professor.

In the US, 1 in 13 children and 1 in 10 adults have a food allergy, she said. And 90 per cent of allergies come from just 8 foods.

For someone eating in a restaurant, the only allergy information normally comes from asking the serving staff, and "they don't always get It right".

LexaTexer

LexaTexer of Berlin is looking at ways to do better analytics on incident reports. The company specialises in unstructured data.

"It isn't possible for a human to read all the reports and they may miss something," said Guenther Hoffman, company founder.

It has already done business with a number of car part suppliers, helping them learn from past incidents.

One idea is to look for data about the environment certain accidents occurred in, such as rain or an earthquake. By spotting trends, safety managers may identify that work in certain environments is too hazardous.

The analytics could also be used to spot patterns, such as certain body parts proving at much higher risk of injury than others, or certain accidents happening at certain times of year, or certain trends in the incidents.

Ultimately it may be possible to correlate someone's level of training with a reduced likelihood of an accident. This would help justify the investment in training, he said.

Invision AI

Invision AI of Toronto (www.invision.ai) is looking at better ways to process data within sensors, so that only important data needs to be communicated, rather than all of it. The jargon is "processing data at the edge", says CEO Karim Ali.

The company is developing a video analytics system which can determine how many participants are in a car, with a live video image of a car travelling at "highway speed". It can be used to enforce regulations where a car must have 2 or more occupants to use a certain lane.

If the entire video stream was sent to the cloud for running analytics, it would take too long – so instead, the analytics is done within the roadside camera.

It has a project with BS Shipmanagement to use video analytics to identify if the oily water separator is being properly operated. This can include facial recognition, to check that the person operating the equipment has authorization.

There is no need to upload any personal data to head office, just data that for example someone is operating equipment without authorization.

Real-time people data from "wearables" – an integrated safety system

Wearables Technologies has developed a software platform to make it easier to handle "wearables" data in an integrated way – such gas, noise and proximity sensors worn by offshore workers

Wearable Technologies Ltd (WTL) of Leicester, UK, has developed a software platform to make it easier for oil and gas companies to handle "wearables" data from their global workforce in an integrated way – such as data from noise, gas and proximity sensors, and people's location.

It is common for offshore workers to carry safety sensors, such as gas detectors, as they work. But until now, this data has only been used in a limited way, to inform the worker themselves, but not the control room.

CEO Mark Bernstein had the idea for the technology after a visit to an oil refinery, where he saw staff carrying portable gas sensors, which could alert the worker to gas, but did not alert co-workers, and didn't alert the control room.

Meanwhile staff in the control room were monitoring equipment but weren't monitoring people. If there was any safety incident, they would not know exactly where people were or their status.

Mr Bernstein's idea was that data about individuals, their location, and the gases they are exposed to, would be managed in an integrated way.

This is similar to the systems which have already been developed to track physical objects in an integrated way.

Coming from a technology background, Mr Bernstein thought it should be a cloud-based wearables "platform", which could gather together data from the various wearable sensors, rather than a specific product.

The technology development was co-funded by BP and is being sold into oil and gas, construction and utilities industries.

There is a question of whether staff will be comfortable being continually monitored and tracked in this way. They should do, so long as they see that it provides a benefit to them, not just the company, Mr Bernstein says. A similar example is the introduction of tachographs in trucks, keeping a record of speed and over time, which were disliked by drivers when first introduced. But now, "tachographs have become the drivers' friend. They prevent unscrupulous employers forcing them to do longer hours. So, this technology can be a force for good," he says.

One common people data platform for North Sea

The company has had a trial running in Aberdeen supported by the Oil and Gas Technology Centre (OGTC), following a May 2018 workshop about "how to make the connected worker a reality" in oil and gas.

The workshop concluded that there "needed to be a common data platform across the industry – that all the oil and gas majors could use, their supply chain, their subcontractors on rigs," he said.

OGTC has funded WTL to upgrade its existing data platform, "So that it could be the standard data platform used across oil and gas in the North Sea."

Partners involved include BP, Wood Group,

Stork, device manufacturer Draeger, and communications supplier Tampnet.

The OGTC funding will cover the cost of developing ATEX certified hardware (which does not give off any spark which could ignite gas). It will also improve the software platform to make it sufficiently scaleable, reliable and secure that it can be easily integrated into large IT systems.

The technology is being pilot tested on a number of oil and gas sites, both onshore and offshore. The data platform is built upon open standards, enabling any other company to integrate their products into it.

How it works

To explain how it works in more detail - offshore workers typically carry a number of devices, monitoring their location (GPS), their exposure to noise and gas and, in some cases, their proximity to moving equipment such as cranes, forklifts and vehicles. They also carry radio communications.

By integrating this data in an automated way and communicating it back to a control centre, supervisors and the control room could be immediately notified if the person is at risk. A real-time holistic picture is also created showing the short term and long term risks each worker has been exposed to, during a shift or even during a whole year.

The WTL technology involves a wearable hub device attached to the worker's clothing which receives, analyses and integrates the data wirelessly from multiple sensing devices carried by the worker. This data is transmitted, including the person's identification where relevant, by the hub device to a company server or cloud system via cellular communication. Workers are identified to the hub using a unique card or ID number, eliminating the risk of data being associated with the wrong person.

The wearable hub also provides power to devices, such as lights embedded in smart PPE garments, where necessary.

If workers are indoors and out of the range of GPS, companies can install Bluetooth beacons on the ground, which enable the system to track when a worker comes in close proximity to them.

The system can also be used to monitor who is on site and where they are.

Mark Bernstein, CEO, has a technology start-up background, with a list of past projects including in virtual reality, a big social network sold to Yahoo, a computer games business which was floated in the UK, and a mobile payments business. He has also been involved in hardware and smart garments projects.

The system is device agnostic – designed to make it possible for customers to add other sensors as and when they choose to, including sensors they already have. This avoids the need for multiple dashboards for each different type of sensor - data from any number of different sensors can be integrated and displayed on a single dashboard.

Having all the data in one place, integrated together in a standard 'data set', makes it easier to do further analytics to get more insights into the overall level of risk different workers have been subjected to, or compare working environments.

Mr Bernstein sees the "connected worker" as a step following the "connected home" and "connected factory", with multiple devices bringing information together.

The company believes it may be serving a $10bn market in the connected worker space, with very few companies in it.

Further uses

The system can be expanded to include many other devices, such as sensors to monitor heartbeat, respiration and heat stress, or even posture, since "back pain is the biggest cause of days off."

The system can send emergency alerts, both to the worker and to the supervisor, if anyone is exposed to gas or noise levels which are dangerously high. "We can monitor the actual noise environment for each worker for months, build up a good long-term occupational health picture.

The data can be used in any legal cases, for example if there is a claim that an employee has damaged hearing due to sustained exposure to high levels of noise. "It is quite difficult to defend if you don't have actual data from that actual employee," he says.

It may be possible to use data analytics, to identify which workers are being placed under the highest levels of risk and from what, which could be useful to insurance companies.

"A customer can decide what they want to monitor," he says. "If a sensor exists, we can integrate it.

Our solution is really an integration and communications platform that sits on top of all those devices."

SecurityScorecard – assessing cybersecurity from outside

SecurityScorecard, of New York City, provides a service to assess a company's cybersecurity without internal access to their computer systems, as a way for companies to assess their partners and suppliers

SecurityScorecard, based in New York City, offers a cyber security assessment of any company in the world, based on externally available information (without internal access to any of the company's systems). The company says it is already used by "an abundance of upstream oil and gas customers" (but they cannot be named).

It is easy to imagine why the service might be useful for oil and gas operators – to assess suppliers before giving them access to parts of the operator's systems, or to assess their risk. It would also be useful to insurance companies, before considering insuring a company against cyber risk. But the big question is, can it work?

If it can work, then cybersecurity ratings might be treated in a similar way to credit ratings. Everybody has a score.

Paul Gagliardi, head of threat intelligence and CISO at SecurityScorecard, says that one of the easiest ways to assess a company's cybersecurity from outside is from records of

Paul Gagliardi, head of threat intelligence and CISO at SecurityScorecard

whether their externally facing website has ever been hacked.

SecurityScorecard has records of 30,000 website breaches since 1998.

From outside, you can scan the website for malware. You can also see how well the website is configured, for example if it is using the latest version of Wordpress.

Another source of information is when companies are obliged to report hacking – such as because they are stock listed, or under the EU's GDPR regulation. Some industry sectors are very active in reporting breaches, such as the US health sector which will often report issues such as a doctor losing a laptop with unencrypted patient data stored on it.

Another source is analysis of the 'end points' – the devices which employees use to access web pages.

One source of data is online advertising companies, which collect data about the browser customers are using, including the version of the browser software, or version of the operating system (both Windows and Android). SecurityScorecard partners with an advertising company, to access its data.

This data can be connected to the IP address of the person operating the device, although not the individual. But SecurityScorecard is using a variety of publicly available data to connect the IP address with the physical building, and the company which uses that account. Typically, a certain physical office for a company will be allocated an IP address, or a range of IP addresses, by its internet service provider, the company says. Connecting the IP address to the company name is a seriously heavy technical lift," he says. "We've invested a lot of resources in that".

If the company is hosting the website itself, then it is possible to see how well patched the server is, or what version of software it is running.

SecurityScorecard also looks at how quickly software updates are implemented – what it calls the "patching cadence".

Other sources of information are "hacker chatter" – if hackers are talking about a company being easy to penetrate.

If a company uses cloud-based systems, it may be possible from outside to determine what systems are being used, and if the company hosting them has a good security score itself.

Another assessment method is spear phishing, sending e-mails designed to trick the recipient into sharing usernames and passwords. These catch out companies which otherwise have very good technical defences. Security Scorecard does not do its own spear phishing tests, but it has found ways to buy domains used in spear phishing tests so it can see who is clicking on the e-mails.

Useful guidance

As well as helping assess clients and partners, the analysis service can help a client better understand its digital 'estate', including elements which are out of the control of the IT department.

Many companies do not have very good co-ordination internally about their cybersecurity, with different elements of it run by different people.

For example, a large company may work with an external provider for its HR, which has lower cybersecurity 'hygiene', but is a close partner with the large company.

Or a public facing website might be produced by a marketing department with no technical expertise, and the IT people, who know about cybersecurity, were not invited to get involved.

SecurityScorecard uses a castle and moat analogy to explain the service to customers. A large company has many partner companies and suppliers which need to sit within its security 'moat'. You have a way of making sure the entire moat can't be breached although you don't have direct control over it.

There have been reports of companies saying they will delist suppliers who get low scores – but since the service is just a guidance, not entirely fact based, it may be better to say, if a company gets a low score it should be subject to more levels of audit of its security, perhaps with IT security staff from the customer doing a physical inspection or phone interview, Mr Gagliardi says.

How it works

Companies are graded A, B, C, D or F. You can receive the score for your own company free by entering your company e-mail address on the SecurityScorecard website.

The score is only an estimate, but its accuracy should increase over time.

Of course there is no certainty that an "A" rated company will never be breached, or that an "F" rated company is very vulnerable to cyberattacks. But it does give an indication of the company's "cyber hygiene", Mr. Gagliardi says. Typically, if a site is breached the company drops down a rating.

Looking at the whole data set, the company can estimate that certain clients are 4-6 times more likely to be breached than others and see that their estimation is true. "We're always tuning that algorithm to make sure that breach likelihood is accurate," he says.

The calculation algorithm is 'normalised' to take into consideration the fact that the bigger a company is, the more that can go wrong. And it is not fair to consider a hack by a highly skilled state as equivalent to hack by a teenager messing around.

But smaller companies might be expected to at least make their website hack-proof.

Hearing that they have a poor score is a strong motivator to companies to include how they manage cybersecurity, Mr Gagliardi says. It is something of a "public embarrassment" which can push people to patch more often.

But some companies may be "a bit combative," the first time the IT manager sees their grade.

Any company has the opportunity to enter a discussion with SecurityScorecard about why they think their grade should be higher.

Oil and gas

Looking at the oil and gas sector, some large oil companies might have tens of thousands of vendors. They may want to receive an alert if any of their vendors have malware on their websites, or other signs of a badly managed cybersecurity, Mr Gagliardi says.

One oil and gas specific concern is the goal conflict between people who want to keep industrial control systems running at all costs, and the desire for cybersecurity, he says.

In traditional IT security, the three pillars of integrity, confidentiality and availability are given equal weighting, but in the operational control systems world, there's much more weighting to "availability".

So, companies often end up using very old versions of Windows and old routers. They don't update them because they are more concerned about keeping the system running, and possible patches causing software problems, than they are of getting a security breach.

But when they don't update the machines and keep them at ancient versions the 'attack surface is significantly increased," he says.

How an E&P saved £500k from better scheduling

An oil and gas operator added £500k to its annual earnings through better scheduling its drilling program, with the help of software from Actenum

US oil and gas software company Actenum reports that it enabled an oil and gas company active in the Eagle Ford shale to add £500k to its annual earnings, through being better able to capture all of its program activities and milestones for a drilling program in a single schedule on the software.

The name of the oil and gas company was undisclosed.

Most of the savings came from the automatic optimisation and scenario evaluation capabilities built into the software, Actenum says.

Each separate optimization run is saved as a scenario and compared to other scenarios, enabling the team to evaluate and select the best one at any point.

Problems

In 2015 it became clear to the operator that, as it had grown by acquiring acreage, it had also outgrown its existing spreadsheet-based well delivery scheduling tool.

Besides being cumbersome and unreliable, tracking all of the operational activities and associated timing and costs was taking more and more time as operational changes piled up.

There was no easy way to evaluate schedule scenarios to determine the impact on capital and production of adding or redeploying rigs and frac crews.

Schedulers had to perform multiple iterations of painstaking manual spreadsheet manipulation to analyze outcomes, while trying to synchronize multiple spreadsheets containing all the schedule data.

Master schedules were kept on a central Microsoft SharePoint server. But downloading and opening the spreadsheet-based schedules was very slow, forcing some planners to rely on colleagues for updates. Or worse, they were looking at outdated information, instead of reviewing the latest schedule version.

Because the tool was spreadsheet-based, the schedulers could not rely on automatic quality control over regulatory stipulations or resource allocation.

Frac and drilling schedules were separate from each other, and from construction schedules, so identifying all resource requirements and timing was difficult, leading to schedule conflicts, poor productivity, and overscheduled resources.

Long-term planning presented a significant challenge because considering all constraints, including permit times, material requirements, and facility construction durations, was simply too time-consuming.

Comparing drilling schedule versions was difficult and laborious. And associating important data with schedules, such as type curves for well production, expected vs actual activity costs, and geographic coordinates for wells, pads, and resources, was not feasible.

Scheduling shortcomings didn't just lead to frustration. They were causing serious red ink in the financial department, Actenum says.

Planners determined that when changes were required as a result of a schedule conflict, an average of two extra rig moves were needed, along with other equipment redeployment. At an approximate cost of $60,000 per move, and an average of four major conflicts each year, $480,000 in overruns were adding to expenses annually.

As well as rigs and completion equipment problems, human resources were also not optimally deployed. The team calculated that at least three people per week were wasting time because of poor scheduling, equating to over $30,000 in unnecessary costs annually.

To staunch financial bleeding and make planning and scheduling more reliable, the operator looked for an integrated scheduling solution that would provide scenario capabilities and enable easy collaboration, as well as predict capital and production impacts by accommodating required data items.

Benefits

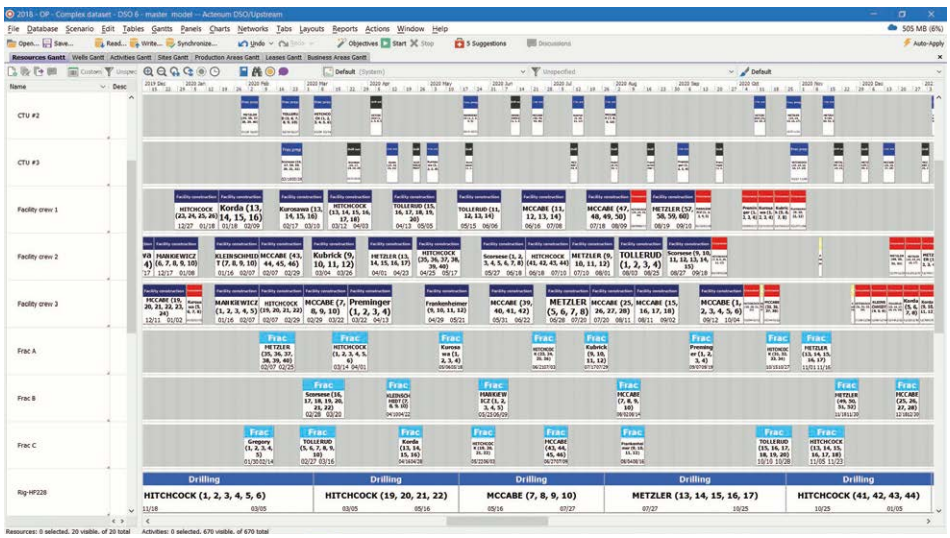
After an extensive evaluation of potential solutions, the operator determined that Actenum's scheduling and optimization solution provided everything that they were looking for, in a feature-loaded, flexible software tool, Actenum says.

The tool incorporates permitting, approval, construction, facilities, drilling, fracking and production activity timing in one integrated schedule, while honoring resource availability and regulatory constraints.

All data relevant to drilling operations is accommodated, and the easy-to-use interface provides drag-and-drop scheduling features. It provides both optimization and built-in scenario analysis to meet the operator's goals to grow and maximize profitability.

The operator has improved planning by incorporating all relevant pre-drill events into the drilling scheduling process. Scenarios include what is feasible to accomplish over the full well lifecycle, not just drilling.

The team is also able to estimate production forecasts accurately by including type well information.



This screenshot shows a resource-based Gantt chart view of a drilling program schedule. Any type of activity and any type of resource may be scheduled.

Cortez Subsea – developments with 3D subsea imagery

Cortez Subsea of Aberdeen is making advances with its technology to generate 3D images of the seabed based on laser scan and optical images - *By Jennifer Green, communications manager, Cortez Subsea*

Cortez Subsea of Aberdeen is making advances with its technology to generate 3D images of the seabed based on laser scan and optical images. It calls it "photo realistic cloud" (PRC) technology.

Scientists say more is known about the surface of the moon than our own ocean floor. But technology in the energy sector advances at an astonishing rate.

Based in Aberdeen, Cortez Subsea was founded in 2011. It is now part of a three-strong group of companies with offices in the UK, Malaysia and Egypt. Its sister companies, DeepTech and MCS, are specialists in remote operations and software.

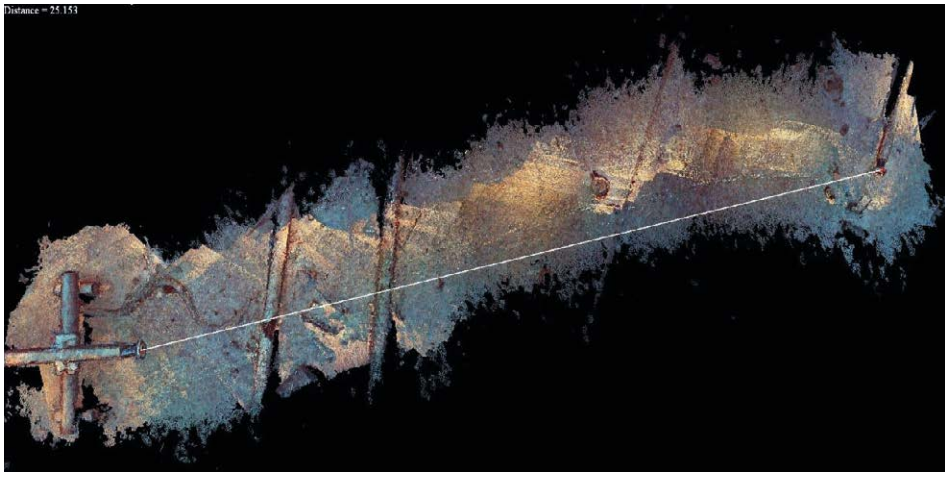
PRC scans structures and pipeline sections underwater in order to create a 3D cloud with billions of points, presenting an as-built, 3D visualisation of any scanned object.

The PRC digital data can be logged and processed by the group's Platform or Pipeline Commander software.

This data can be shared on and offshore and in real-time.

"Conventional reporting can take around four weeks. NuWave reporting is fast, with a draft report compiled before the team depart and final reports produced within 14 days," says Alasdair Cowie, managing director of Cortez Subsea.

"As data is accumulated over subsequent inspections, precision comparison can take place to assess changes and plan pro-active maintenance."



A photo-realistic cloud (PRC) for measurement of the closing spool, connecting new wells to existing plant

The technology can assist in applications such as 'out of straightness survey', pipeline lateral and vertical buckling surveys, spool measurements, subsea construction, decommissioning, repairs, chain inspection, FPSO inspection and onshore 3D modelling using drones.

The PRC can "overcome operational hurdles" where traditional methods "fail or have become obsolete," Mr Cowie says.

The technology was used in the Gulf of Suez, offshore Egypt, to carry out the installation of nine subsea spool pieces at a range of water depths between 15m and 75m.

The project included closing spool metrology (measurement), crossing and clearance metrology, proposed route survey, as built survey, and free span measurements.

Two techniques were deployed from a dive support vessel - taut wire metrology using a diver and PRC technology deployed by an ROV (remote operated vessel). This allowed clear comparisons to be drawn between the techniques and highlighted both the drawbacks of using divers in this scenario and the advantages of the PRC technology.

"Using an ROV and the PRC for the spool metrology significantly reduced the time required for this activity," says Mr Cowie. "15 minutes for the ROV to fly and video the route of the spool from the pipeline to the riser using the PRC, in comparison to two hours required for a diver to perform the traditional taut wire process."

The data from the PRC was transmitted to the surface in real-time.

"The PRC system operates in good to poor visibility. On this occasion there was good visibility at the seabed and therefore this allowed a very high level of accuracy of 1/1000 per meter to be achieved.

"This far exceeds the traditional taut wire solution whereby the measured lengths and angles were not accurate because of wire sagging and diver's estimations."

"The ROV and PRC were able to survey areas in the vicinity of the platform facility with multiple pipeline crossings and obstacle clashes, whereas the divers could not pass or maneuver freely around these due to diving operability.

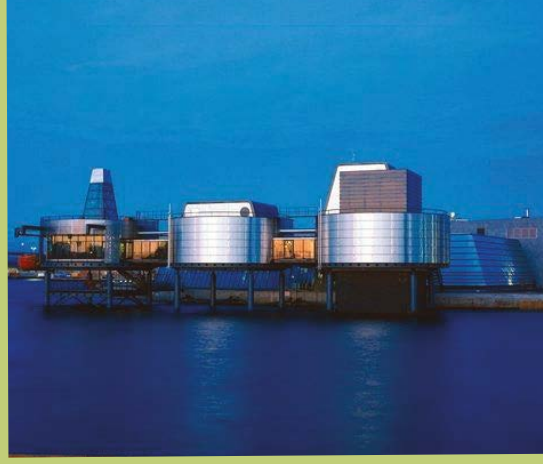
"The 3D model produced for this metrology serves as an as-built model of the subsea installation and the surrounding area which provides a baseline for future intervention and inspection works."

"NuWave offers dive-free, high-tech inspection of platforms and pipelines." Mr Cowie says.

As built survey – subsea imagery of what has been built

Finding Petroleum

digital energy journal



Understanding better ways to work with technology to meet business goals

Events 2020

Opportunities in NW Europe
Maintain production and explore new technology -
London, 23 Mar 2020

Fractured Reservoirs: Examples & New Geophysics
Advances in seismic technolgy that enable us to become entirely predictive about fractures!
London, 27 Apr 2020

Finding Oil and Gas in Southern Africa
The hotspots across Southern Africa - offshore South Africa, Mocambique, Namibia and Angola

Special free event at South Africa High Commission in London

London, 5 May 2020

Finding Petroleum in the Middle East
Where are the opportunities for investors and E&Ps in the Middle East?
London, 11 Jun 2020

Opportunities in the Eastern Mediterranean
Greece, Italy, Cyprus, Lebanon
London, 25 Sep 2020

Digital Ship Malaysia - How to Digitalise Operations and Wells
Kuala Lumpur, 6 - 8 Oct 2020

How to Digitalise Exploration
Where are the most interesting digital technology advances?
London, 11 Nov 2020

Carbon Capture and ESG
What satisfies ESG investors - and makes a business work for E&Ps?
London, 10 Dec 2020

Interested in speaking / sponsoring a Finding Petroleum event in 2020?
Download our guide to marketers at http://bit.ly/FPMP2020

admission from £50

Find out more and reserve your place at

www.d-e-j.com

www.findingpetroleum.com