



PPDM™

# Foundations

Journal of the Professional Petroleum Data Management Association

Print: ISSN 2368-7533 - Online: ISSN 2368-7541

Volume 5 | Issue 2



**1st Place** Foundations Photo Contest Winner; *"Offshore Drilling in Malaysia"* (Jim Boud)

## Geophysical Data Compliancy

Utilizing Technology (Page 4)

### PLUS PHOTO CONTEST:

This issues winners and how to enter (Page 19)

# Make Smarter Decisions

EnergyIQ creates information certainty  
in an uncertain E&P world.



## **IQlifecycle™**

Be more productive and efficient by automating business processes across the life of the well with the leading well master solution.



## **IQgeoscience™**

Create comprehensive, current, and quality data and give your team the confidence they need to make better decisions faster.



## **IQinsights™**

Stop your search and start your discovery - instantly discover information to make rapid and accurate analytical decisions with certainty.

EnergyIQ's E&P Master Data Management solutions are used by companies that want to build better wells, streamline processes, create and access better data, and live in a data-driven world.

**303.790.0919**

[EnergyIQ.info](http://EnergyIQ.info) • [info@EnergyIQ.info](mailto:info@EnergyIQ.info)

© 2018 EnergyIQ, LLC. All Rights Reserved.

**energyIQ**  
Information Certainty

# Foundations

Foundations: The Journal of the Professional  
Petroleum Data Management Association.



PPDM™

## Chief Executive Officer

Trudy Curtis

## Senior Operations Coordinator

Amanda Phillips

## Senior Community Development Coordinator

Elise Sommer

## Article Contributors/Authors

Amii Bean, Sue Carr, Jim Crompton,  
Dave Fisher, Tracy Heim, Guy Holmes,  
Jess Kozman, Marina Melekhina, Trish  
Mulder, Ramon Perdomo, Charity Queret,  
Scott Shemwell, Hazel Welch

## Editorial Assistance

Courtney Cooper, Beci Carrington,  
Jim Crompton, Dave Fisher

## Graphics & Illustrations

Jasleen Virdi  
Freepik

## Graphic Design



## BOARD OF DIRECTORS

### Chair

Robert Best

### Vice Chair

Allan Huber

### Secretary

Lesley Evans

### Treasurer

Peter MacDougall

### Directors

Allan Huber, Ali Sangster, Amii Bean,  
Christine Miesner, David Hood, Jeremy  
Eade, Lesley Evans, Paloma Urbano,  
Peter MacDougall, Robert Best, Shashank  
Panchangam, Trevor Hicks

### Head Office

Suite 860, 736 8th Ave SW  
Calgary, AB T2P 1H4, Canada  
Email: info@ppdm.org  
Phone: 1-403-660-7817

# Table of Contents

Volume 5 | Issue 2

## COVER FEATURE

### Geophysical Data Compliancy

4

Utilizing Technology.

By Sue Carr & Trish Mulder

## GUEST EDITORIALS

### World First At Houston Expo

10

The Cloud.

By Guy Holmes

### An Update On Digital Literacy

12

The New Workforce.

By Jim Crompton

### Ethics of Big Data

20

And Optimization Tools.

By Ramon Perdomo &  
Marina Melekhina

### The Value Of A CPDA

22

Worth The Investment.

By Hazel Welch

### Six Data Testing Practices

32

To Increase Business Intelligence ROI.

By Charity Quaret

## FEATURES

### How To Make The Case

6

To The C-Suite.

By Scott Shemwell

### What Is A Completion?

8

An Update.

By Dave Fisher

## Infographic

16

PPDM's Year In Review.

By PPDM

## Two Database Giants

24

SQL And NoSQL.

By Amii Bean

## Thank You To Our Volunteers

26

Featuring Three Great Volunteers.

## Hands On With The PPDM Association Board Of Directors

27

To SQL Or NoSQL: That Is The Question.

By Amii Bean

## Petroleum Data Management

28

As A Professional Discipline.

By Tracy Heim

## DEPARTMENTS

### Photo Contest

19

This issue's winners and how  
YOU can get your photo on the cover  
of Foundations.

### Upcoming Events, Training and Certification

35

Join PPDM at events and conferences  
around the world in 2018 and  
2019. Learn about upcoming CPDA  
Examination dates and training  
opportunities.

## ABOUT PPDM

The Professional Petroleum Data Management Association (PPDM) is the not for profit, global society that enables the development of professional data managers, engages them in community, and endorses a collective body of knowledge for data management across the oil and gas industry.

Publish Date: October 2018



# Geophysical Data Compliance – Utilizing Technology

By Sue Carr & Trish Mulder, Katalyst Data Management

**G**eophysical data compliance means only using data you are entitled to use for an implicitly defined purpose and timeframe. Geophysical / seismic data is complex. Seismic is an image of a defined position of the earth. It is multi-dimensional, multi-faceted and geographically based. It consists of raw and interpreted data and could reside on different media types and formats. It may be stored internally or at an external storage facility. Along with this, the Oil and Gas industry is an ever changing and complex environment due to mergers and acquisitions, contractual laws and obligations either to government agencies, exploration companies or multi-spec corporations. As artificial intelligence and big data define a significant path through the information management world to allow corporations to make more informed and educated decisions, so too must our data management domain. It must adapt and modernize itself to keep up with the demand for information and ensure that it

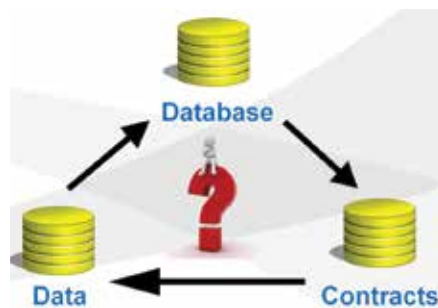
meets its legal rights and ethical obligations. Validating such data and information in ones corporate environment can seem like a monumental endeavor. However, technology can contribute to greater success of projects by sifting through large volumes of data, providing data analytics and constantly monitoring the data environment to ensure that your company remains data compliant.

One model breaks down the overall compliance progression into a six phase process which continuously cycles into a data governance atmosphere. The six phases consist of three asset phases and three transitional phases. All phases must be completed, and may occur simultaneously; in order to achieve compliance. This model is different from others as it recognizes a corporation can achieve compliance by relying on technology, visualization tools and repeatable monitoring systems. This article is the first in a two part series, and will focus on the three asset phases (the database, contracts and

the data). The second will complete the overall method with a discussion of the three transitional phases.

The Database is the first asset phase, where the data compliance process begins and ultimately allows a corporation to move into a long term data governance environment. It is your single point of access and source of truth. As the data management domain adapts and modernizes, so too are the applications to which it utilizes. Given the trends and shift towards artificial intelligence, software developers are recognizing and understanding the importance of data analytics and are evolving their data management suites to include more dashboarding and reporting capabilities, together with better visualization functionality.

It is here that analytical tools must first be incorporated against your database. If your application currently does not have these analytical tools integrated, you will need to run your database against other external sources. There are a



plethora of cheap options at your disposal. A discussion with your IT department will assist in selecting an analytical tool that will work within the corporate environment. This step will be critical in the latter phases of the compliance process.

The second phase focusses on the Contracts, which is your source of entitlements. The critical component in this phase is to locate all contracts. This, of course, is easier said than done, as contracts are usually managed within different departments such as accounting, land, legal, earth sciences, and the records team. A clear understanding of internal department record taxonomies and access levels (even if only read access) to their documents is crucial for this phase to be successful. With assistance from your IT department, your company can unleash crawling technology within your internal environment. Once established, determine specific business rules for what the crawler is to look for. Every company has an infinite number of documents that the crawler needs to sift through, so the more defined business rules you create, the fewer documents or locations the program has to crawl through. This will help to reduce works scope and produce faster and more meaningful results.

Once contracts have been located, all physical or previously scanned documents should be processed through an Optical Character Recognition (OCR) application. This will enhance your ability to extract information automatically from the documents for easy population into your database and reduce some of the traditionally labor intensive methods. There is a challenge. Even though OCR tools have made great strides over the

past decade and there are many to select from, not all OCR applications are created equal and therefore you must do proper due diligence prior to selecting a program. The purpose of an OCR tool is to extract information from the document in order to minimize the typical manual workflow. The system will require upfront “training” for it to understand the type of documents it is examining and what information to extract. More training will allow the tool to become faster at recognizing key differences. Again, programs such as these are merely a mechanism to assist and reduce time and manual effort.

The last phase is the Data. This phase consists of two segments, (1) data stored externally and (2) data stored internally, often called the user environment. This is where the Geophysicist will interpret the data. Each is handled in the same manner, however internal data normally experiences more challenges than external given that (1) a level of participation is now required from the user community and (2) data filenames historically can, and have, been manipulated.

This phase is where your entire corporation’s effort comes together with the purpose to comprehend the big picture. Utilizing a GIS/visualization program to display the information gathered is a great and easy way for people to understand the sheer complexity, severity and liability of your internal data environment. It also provides a clear end path to achieving data compliance. Tools can be scripted and used to interrogate the contents of an interpretation project and the SEGY files themselves in order to ‘scrape’ the metadata from the binary, EBCDIC and trace headers to find unique identifiers. This information can then be matched to your source of truth (database) and source of entitlement (contracts) and, when compared to the navigational data, can provide a level of confidence that the data is assigned to the correct survey in the database and contract. User intervention and participation is imperative as reporting mechanisms can only identify the likelihood that something

is a match. The information provided within the report must be validated and either accepted or rejected. Once accepted, the data will be manipulated and linked so that the data, database and contract can operate as an integrated asset.

Becoming a data compliant corporation is no small feat and active participation from many departments is required in order for success. Applying technology wherever possible allows you to see the complexity of the situation and in turn can minimize labor intensive processes. Upfront effort and costs may be required, and the long term gain and sustainability of your corporation’s data governance achievements will rely heavily on the consistent monitoring and reporting you establish today. Changes to information acts, copyright laws and professional standards are making data compliance critical and imperative that corporations are taking the right steps to ensure that data used meets strict contractual obligations. ■

#### About the Authors

Sue Carr, Manager Consulting Services, Katalyst DM: over 35 years of implementing software and data management systems and leading subsurface data teams. Sue is focused on building a DM Consultants Group to help solve E&P companies’ data challenges.

Trish Mulder, Director; Business Development, Katalyst DM: a seismic data expert with over 18 years of experience in data management at both E&P and service companies. Trish has a strong vision for the future of data transactions and compliance.



# How To Make The Case To The C-Suite

## Selling Large Scale Data Management Projects to the C-Suite

By Scott Shemwell, The Rapid Response Institute

**M**ichelle H., the Chief Data Scientist for a large independent E&P company very active in shale plays, returned from a conference focused on the Future of the Upstream Digital Enterprise. All speakers and panelists extolled the enterprise digitalization process, the new role of data management and emerging game changing impact.

With the explosion in data quantities, types and new tools, Michelle suggested to Aaron, the CIO, that the company adopt an aggressive strategy to fully implement the company's digital oilfield initiative, which incorporates the PPDM Data Model. Initially, Aaron did not share Michelle's enthusiasm.

Since 2015, the company had undergone several restructuring processes and the CFO is still very concerned about cash flow. Moreover, he, the CEO and Board share the belief that the current WTI crude oil price is not sustainable and that \$50 or less is more likely than \$80 per bbl.

When Aaron asked Michelle about the risk of this undertaking, her response was not reassuring. At the conference, major consulting firms and other "experts" had explicitly claimed that organizations had no choice but to digitalize. The subject of project risk was only briefly mentioned.

Aaron asked Michelle and her team to develop a project plan, capital budget, operational costs and risk mitigation model. Finally, he wanted to know if the highest value to the firm was as an industry leader, a fast follower or

even a laggard. Only then would he be comfortable presenting to the CFO.

### PREPARING THE BUSINESS CASE

Aaron has advised Michelle that standard IT Use Case initiatives are perceived as weak, i.e., unsupported statements regarding increased efficiency and productivity, etc.

Perhaps a daunting challenge for Michelle and her team; it need not be. Moreover, the company has developed a set of CAPEX criteria, i.e., alignment with business objectives, return on investment (often measured as Return on Investment Capital [ROIC]), validation, project management, risks, etc. Her team will need to develop a solid understanding of the Capital Review Process and make sure that key questions, concerns and issues are addressed.

She will also need to add an executive sponsor/coach. This should be an individual with credibility with top management. Logically, she may seek the CIO, however, in some organizations this individual may not have the perceived credentials. A senior executive with profit/loss responsibility might be preferred.

Finally, she must be able to address all questions, including ones that may not be anticipated. Should a question be posed that she cannot immediately answer she should make sure she understands the query, doesn't 'wing it' and states emphatically that she will respond in writing by X date.

### QUESTIONS MANAGEMENT MAY ASK

Challenges from management Michelle will need to address may include:

1. *Is the company culture ready? If not, how can we make it ready?*  
While this may appear to be beyond the scope of the Chief Data Scientist, the answer is critical to her success. She will need an in-depth understanding of the organization's culture and have a transformational plan.
2. *Other enterprise level IT initiatives have 'over-promised and underdelivered.' What makes this one different?*  
The answer to this and the first question requires an understanding of where the organization is on the maturity and technology adoption curve. Trying to force a laggard into a leading position will likely end in failure. Moreover, statements to the effect that 'everyone is doing it' will not exude confidence.
3. *How will the organizational Governance model incorporate digitalization as 'the way we run the business?'*

The consulting firm, i-SCOOP defines Digitalization as "Digital transformation is the profound and accelerating transformation of business activities, processes, competencies and models to fully leverage the changes and opportunities of digital technologies and their impact across society in a strategic and prioritized way,



with present and future shifts in mind.” Issues around this disruptive transformation must be incorporated into the initiative. Often IT initiatives have not addressed governance to their detriment.

4. *Technology is advancing rapidly, how do we know this is the right time and technology suite?*

Without going into the technical details, Michelle will have to outline a Technical Architecture that will accommodate the existing legacy systems, current technologies and future technologies.

5. *You're asking for a substantial amount of funding with a multi-year implementation plan. How does the organization mitigate this project risk?*

The project team will be led by experienced project managers. They will take full advantage of project management processes, i.e., Project Management Institute's PMBOK® Guide and Standards.

6. *How can you assure organizational 'buy-in'—the Change Management strategy?*

Change Management is fundamental to the digitalization transformation. It is well understood; however, often not followed. Persistence is necessary, not just within the organization but including its ecosystem.

7. *We hear a lot about Cyber Security. How can you assure that we are not putting the organization at risk if we go forward with this project?*

This hard question must be adequately answered. Multiple hacks on organizations have destroyed shareholder value, ruined careers and

threatened critical infrastructure. Outside expertise may be appropriate in this fast-changing environment.

**ALIGNMENT WITH THE BUSINESS**

It is surprising how many senior executives, especially in support areas such as IT, do not understand the organization's vision, mission and objectives. This is easily remedied by reading the Annual Report, particularly the opening Letter to Shareholders where this information should be clearly stated.

Large CAPEX that are not aligned with the business objectives are non-starters and a waste of everyone's time and resources. Understanding this is fundamental.

Since an enterprise level project will impact all aspect of the business's ecosystem, Michelle will need to develop a detailed assessment of the positive and negative impacts, i.e., time commitments, resource constraints, etc.

**ECONOMIC VALUE PROPOSITION**

Macro statements like “digitalization could impact the bottom line by 30 percent” are great in white papers and at conferences, but management usually cannot get their minds around such aggressive claims. She should refrain from such glittering generalities.

Economic models are challenged to incorporate an understanding of which parts of an organization, including its ecosystem, that are impacted as well as why and how. As noted, statements often made, words to the effect of, “we can save xx percent or increase (blank)” are meaningless to executives.

However, if Michelle assesses the impact on major departments and/or other entities she will accomplish two major objectives. Working with department leadership her team will develop an understanding regarding their needs and value proposition. She will also generate ‘buy-in’ throughout the ecosystem.

Hypothetically, one group may see a one-to-three percent impact while another none. When added together across the ecosystem the final number may in fact

be 30+ percent. Once this exercise is completed, Michelle will know how the organization will receive this initiative.

The 80-20 rule applies. It is not critical to uncover every area of value, just the major impacts.

Finally, every project needs a measurable return and set of KPIs and CFSs. Without robust yet straightforward models, projects are often rejected.

**HUMAN SYSTEMS INTEGRATION**

*Source: The Rapid Response Institute*



Michelle should be prepared to address any Human Systems Integrations concerns raised by management – where the human decision makers meet the interface. For example, the current fervor for the driverless automobile is seen by some as fait accompli. However, what exposure does the firm have if the ‘machines’ make the wrong decision?

**RISK MITIGATION**

One of the most important functions of management is risk mitigation. Historically, the focus has been on financial transparency, managerial ethics, etc, now it includes more concerns like digital safety.

With potentially tens of thousands of sensor exposure access points, as well as major integrated tasks from ecosystem partners and suppliers, traditional IT governance rules no longer apply. Emerging organizational governance models incorporate IT fully with other risk management processes.

Michelle and her team will need to explain to management in business terms

(no geek speak) what the risk mitigation process is for both the project as well as ongoing operations. This is an area where her executive sponsor can assist.

## THE DECISION PROCESS

Once Michelle and Aaron have developed a detailed and supported understanding of the points herein, they are ready to face the CFO, CEO and possibly the Board. The value to the business will be well founded and 'objections' acknowledged and addressed.

In preparation, she must reduce what should be a mountain of information into the 15 minutes the CFO will give her, including her tone setting first impression. Even in an era of casual dress her team must 'dress for success.' Many C level individuals still wear suits and showing up in jeans may dilute her message.

Her presentation must be succinct and address the opportunity, risks and reason why the investment should be made. If she is prepared, her confidence and ability to present her case will likely end in approval. ■

### About the Author

Dr. Scott M. Shemwell,  
sshemwell@theRRInstitute.com,  
Managing Director of The Rapid Response Institute is an acknowledged authority and thought leader in field operations and risk management. He has more than 30 years in the energy sector leading turnaround and transformation processes for global S&P 500 organizations as well as start-up and professional service firms.



# What Is A Completion?

By Dave Fisher,  
Dave Fisher Consulting

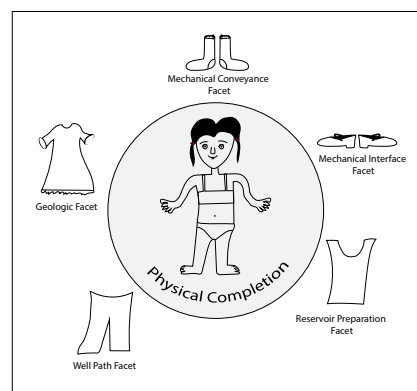
**T**he *What Is A Completion* work group is nearing the end of its project. The full results will be published for our members in the Fall of 2018. Here are some highlights of the work.

If you ask a variety of people in the Oil and Gas industry, "What is a completion?", you are likely to get a variety of answers. Each person has their own understanding of the word, based on how they engage with the thing they call a completion. For example, a completions engineer cares about equipment and fluid flow whereas a business analyst focuses on the financial performance of the asset.

When a data analyst gets a request for "all the information on the completions in this well," what is required? When an operator delivers the completions data to the well's partners, does each partner know how to capture the data elements correctly and completely? Data exchange requires clarity of meaning.

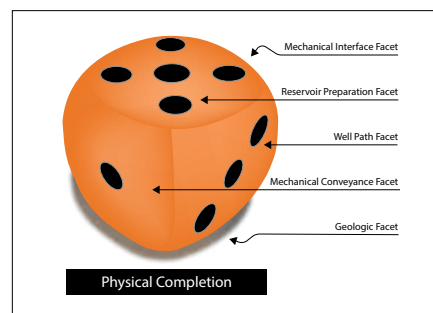
The different stakeholder perspectives are *facets* of the object. Imagine that a paper doll represents a completion object in a well; each child makes a clothing choice. One child wants all the clothes; another doesn't want the socks; another one would be upset if the socks were lost.

The word "completion" also has a very common non-technical definition: the conclusion of an activity. It might mean that the operations at a well have reached a milestone. It might also mean the end of a contract or process that has nothing to do with a well. The context controls the meaning of



the word. If you don't understand the context from the other's viewpoint ("facet"), you may lose the meaning. In the world of paper dolls, if someone starts talking about GI Joe or hunting for Dall sheep in Alaska, the conversation is suddenly on quite a different track.

Facets can also be visualized as faces on a cube. Each face is a different view of the same object.



The PPDM work group for *What Is A Completion* (WIAC) addressed the semantic challenge: how can we facilitate clarity in the industry's use of the word "completion"? It would be impossible to produce one

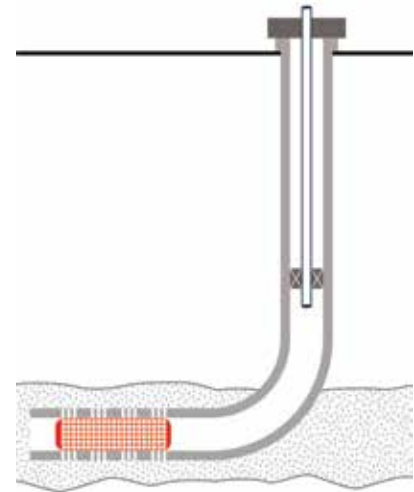


Figure 2

Mechanical Interface Facet: equipment that controls the flow of fluids at the contact interval(s) between the reservoir(s) and the production conveyance mechanism.

**What is a Completion?**

The Professional Petroleum Data Management (PPDM) Association is the not for profit, global society that enables the development of professional data managers, engages them in community, and endorses a collective body of knowledge for data management across the oil and gas industry.

bp

KATALYST

geoLOGIC SYSTEMS

R.1.0 August 2018  
© Copyright 2018, PPDM Association. All Rights Reserved

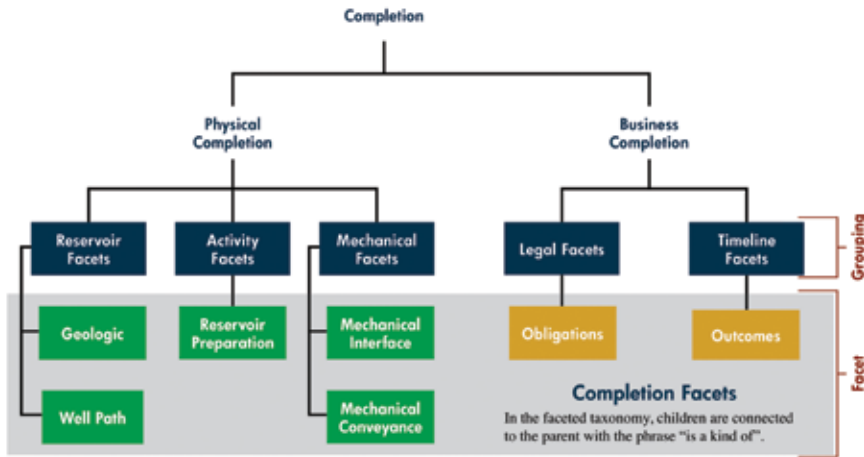


Figure 1

definition that would satisfy everyone’s needs; multiple and conflicting definitions are too widely entrenched. Therefore, the solution is to provide the means for everyone to talk with each other using the common ground of a set of baseline definitions and illustrations. A baseline is a starting point from which variations can be explored and explained. A similar method is in PPDM’s *What Is A Well*.

The work group developed the concept of “completion” related to a physical completion (“a thing or activity in the well”) or a business completion (“something about a well”). Each of these has facets representing the information requirements of the stakeholders. In the diagram, the facet names are shown in Figure 1 in red. Please note that the facets are not necessarily “objects” to be incorporated into a database; their main purpose is simply to facilitate communication between people who want to be sure they understand each other.

The published WIAC product will define the terms and provide supporting illustrations and charts, such as the above. It is a simple representation of the mechanical interface facet (highlighted in red on Figure 2.)

As with other products of the PPDM Association, *What Is A Completion* is

the result of many hours of work by our participating work group volunteers and many reviewers globally. Thank you to all of you for volunteering your time and expertise. Of course none of this would have been possible without the generous support of our Champion Sponsors’ contribution. Thank you! 🙏

**About the Author**

Dave Fisher is a retired Calgary geologist. He had never heard of data models until he joined the PPDM board of directors in 1993.

**THANK YOU TO OUR SPONSORS**



# World First At PPDM's Houston Professional Petroleum Data Expo



By Guy Holmes, Tape Ark

**T**he benefits of the public cloud were highlighted at the PPDM Professional Petroleum Data Expo in Houston on Tuesday, April 17, and Wednesday, April 18, 2018, with a demonstration that involved live seismic acquisition directly into the cloud, with data processing, interpretation and data management being done in one continuous workflow across multiple continents simultaneously. During the presentation, data was shot in Oklahoma and transferred directly into Amazon S3, where automated tools captured metadata, populated a PPDM database and updated a GIS system. The data was then used by a geophysicist in Australia to perform interpretation and other data management functions and once complete, the data was archived in Amazon Glacier for long term retention. The workflow, which is usually performed over several months, and which uses the transport of tapes or hard disks to move the data around, was completed live on stage in the space of 30 minutes. It was a show of how interconnectivity and imagination can lead to thought provoking moments of “Ah Haa.”

When I decided to perform this demonstration, I did so with three key issues in mind that I felt the data management and PPDM community

needed to be made aware of. The demonstration, which involved several commercial entities, was not designed to showcase a particular solution or service, but more to show the data management community what was possible and to push the envelope when it comes to new ways that data management can be handled.

One of the three areas of focus in this demonstration involved how public cloud added massive flexibility to just about any data management solution that is being developed or used when compared to closed private cloud data management solutions. The second area of focus was to show that not only is the technology of today ready to change the industry, but the speed of technological change will see the data management function advance faster than at any other time in history. Lastly, I wanted to highlight that if you want things to change, you have to push the boundaries – even if it is in the form of a risky live demo.

### **PUBLIC VS PRIVATE CLOUD**

Before I can explain why public cloud is, and will be, so important to the oil sector moving forward, I need to explain the major differences between public vs private cloud offerings.

Public Cloud – A public cloud is one

based on a standard cloud computing model, in which a provider (ie. Amazon, Microsoft, Google) can make resources such as virtual machines (VMs), applications, storage, etc. available to the general public over the internet. Public cloud services may be free or offered on a pay-per-usage model. Public cloud is ubiquitous (as a general rule), and due to the highly available nature of it, allows for the development of new technology and workflows to come from anyone who uses it – even if from outside the oil industry. Public cloud also offers the scale and flexibility that private cloud cannot, and this means that private cloud tends to be restrictive in terms of the data storage capacity that can be consumed, the data processing power (CPU) that can be used and its capacity for collaboration. Due to the scalability and ubiquitous nature of public cloud, it also tends to be a far cheaper total cost of ownership – sometimes by orders of magnitude. As an example of the scale and flexibility of public cloud, Netflix alone is responsible for almost 40 percent of the internet traffic on a Sunday night in the United States.

Private Cloud – A private cloud is a particular model of cloud that involves a distinct environment in which only the specified client can operate. Private cloud tends to be geographically restrictive



Guy Holmes presenting at the 2018 Houston Expo

and is often used by companies who need solutions tailored to very specific needs. The scale and flexibility of the private cloud is restricted and often known up front by those who consume it. The ability to collaborate with third parties within a private cloud is usually not possible or is very difficult, and the ecosystem of solutions that can be developed and deployed in the private cloud is almost non-existent unless done on a bespoke basis.

### WHY IS THE PUBLIC CLOUD TAKING OVER OR ELIMINATING PRIVATE CLOUD?

Companies that offer private cloud solutions tend to work on the basis of solving one issue for a client, like pure storage of data. The reality is that the needs of oil companies are far too complex and the ecosystem of potential suppliers of technology is growing at such a rapid rate that having your company's data locked in a private cloud where none of the new technology can access it, is no longer acceptable. In the next 10 years oil companies will be serviced by a much wider audience of service providers, app developers, analytics providers and data processing systems. Putting data into private clouds will prevent companies from getting access to the new wave of vendors and toolsets that have been developed already or will be developed in the future. We have seen examples of machine learning and artificial intelligence systems being run on Oil and Gas data to provide

reserve estimates, drilling optimisation and subsurface analytics that are going to markedly change the way we work in the industry. Pouring data into a private cloud will likely mean that access to this new technology will either be technically very challenging, cost prohibitive or simply not scalable enough to withstand the Oil and Gas style of workload.

There are no private clouds that I am aware of that could sustain the volumes of data from major oil companies or provide the scalability of processing capability and CPU volumes, needed to process the data. The reality is that public clouds are now capable of having an entire oil company's data collection online and accessible, which was only just a dream a few years ago.


### SPEED OF TECHNOLOGY DEVELOPMENT

The speed at which new technology is being developed is faster than at any time in history. It is broadly acknowledged that machine learning, artificial intelligence and advances in big data and analytics stand to make most significant changes and technological advances in the oil sector over the next 10 years. New ways to explore, new ways to confidently identify targets and new ways to manage supply and demand, maintenance, and safety are all on the visible horizon. Technology developments that take place in disparate industries will start to be cross pollinated to fill new needs. A pipeline company whose assets suffer from occasional pipeline blockages or valve failures, will

benefit from technology developed by medical research in the field of cardiology on heart valves and arteries. Space research in the use of robots that help build colonies on other celestial bodies, will be used to help with the needs of the construction of assets on the sea floor used for the drilling industry.

The pace of technological evolution seems to already be outpacing the ability of consumers to adopt or even understand the technology before the next generation of technology has been released. The oil industry, which has typically lagged in the adoption of new technology in the IT space, and once dominated by large multinational suppliers, will become fragmented and diverse in both offering and uniqueness.

### PUSHING BOUNDARIES

Back to a world first at the PPDM Houston Expo. The purpose of the demonstration I did at the Expo was not to sell a solution, but to showcase what can be done in automated, multi-geographic data acquisition and processing. I did it to show how the adherence to standards in the industry can facilitate data transfer, metadata capture, and data sharing. I did it to show how the PPDM data model can be used to share data between applications in real time. I did it to show how the speed of access to data can change the way we perform our surveys, QC our data and dynamically interact with our service providers. I did it to get data managers to think outside the box and to start thinking about what is possible rather than how have we always done it. The Oil and Gas industry is now in a position to pivot into new technology frontiers and in some ways catch-up with the rest of the world when it comes to adopting new and lean ways of working. I did it because I could. 

#### About the Author

Over the past 19 years Guy has chased his passions wherever they led. In some cases, his passion led him to starting a company that imported wine accessories, and in another he founded a leading global data management company.



# An Update on Digital Literacy

By Jim Crompton, Reflections Data Consulting

**A**ll the headlines seem to be about technology. From AI (Artificial Intelligence) to ML (Machine Learning) to analytics platforms, to robotics and automation (don't forget drones), to Big Data and open source platforms like Hadoop and Spark, we can't get enough of the new toys. But I want to return to one of my favorite topics, what about humans in this world of emerging digital technology advances?

Here is a quick quote from a recent Economist article (January 14, 2017): "When education fails to keep pace with technology, the result is inequality. Without the skills to stay useful as innovations arrive, workers suffer – and if enough of them fall behind, society starts to fall apart." I don't want to sound the alarm for societal collapse here but keeping a sharp eye out for digital literacy seems to be a sensible thing for employees and employers.

Digital Literacy can be a broad topic, so in this article I just want to touch on five subjects: how we interact with computers, how we interact with data, how we use technology to collaborate, are we learning appropriate lessons from social media and gaming, and finally the difficult topic of AI versus HI (Human Intelligence, expertise and experience).

The human computer interface is undergoing an amazing transformation. For years, programmers tried to make computers easier to work with from

operating systems to applications, but all it seemed to mean was that our technology needed "system friendly users" rather than being a "user friendly system." So, we trained and some people got it, but most people struggled.

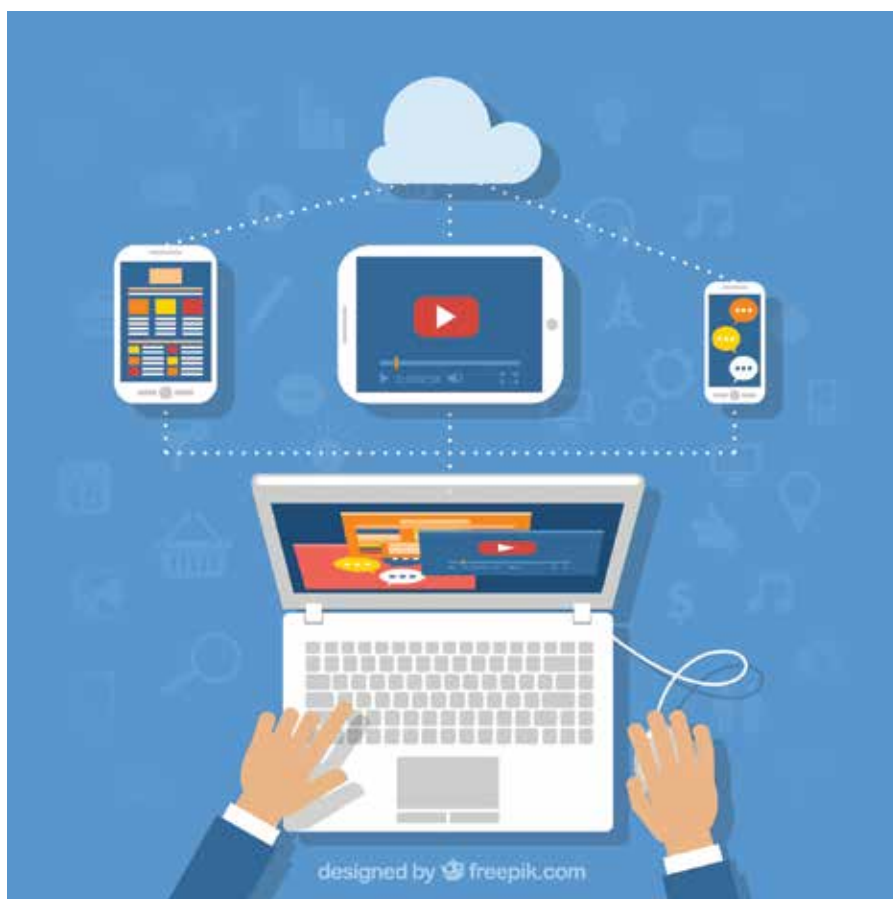
Now comes a new interface paradigm, voice. Speech recognition and machine translation have made remarkable advances. Computerized personal assistants, such as Apple's Siri, Amazon's Alexa, Google Now and Microsoft's Cortana, can now take a wide variety of questions and return accurate and useful answers in a friendly human sounding voice. So, while more intuitive keyboard and website interfaces and touch screen devices have added value, the future maybe talking to our technology will be the best approach for many of us.

The practical definition of Big Data is a condition where we have more data than we can handle. As humans that threshold comes pretty low. The way we combat this condition is from better techniques with data visualization, data discovery and data mining. The art of data visualization is not new but better tools from Spotfire, Tableau and Qlik are now available to help folks understand what data is available and how to make sense of it. Visualization is an important link between what the computer does best (data processing) and what the human does best (conceptualization).

Digital literacy isn't just about working with computers. There is great opportunity

in using digital technology to help us work with other humans, the collaboration platform. With the growing use of integrated operations centers, decision support centers (or whatever the operator or service company calls it), technology can facilitate humans in different places (offshore versus onshore, experts in different locations and supply chain partners in different offices) to reach better decision faster. The goals can be from basic surveillance (what is going on), to analysis (why is it doing that) to predictive and prescriptive steps (I want to fix it before it breaks and keep it working that way). Collaboration is an important role of digital literacy and it requires learning some new languages (*What is a Well*, etc.).

The new "digital native" is slowly filling the workforce as the "digital luddites" (like me) depart. These new employees are natural networkers and highly skilled at leveraging digital technology. But that skill is so far limited to sharing vacation photos and arranging to get together for a pizza and a beer on Friday, or how to best kill alien monsters and capture the hidden treasure. Learning how to turn the lessons of virtual messaging, virtual reality and even augmented reality technologies to the tasks in the oilfield can be done, but requires a different way of thinking. As well as teaching the new workforce the way we have always done things, we need to listen to them for suggestions on how to do things better.



Some believe (and I am not in this camp) that human experience and expertise can all be built into advanced algorithms and turned over (through automation) to technology to run the oilfield of the future. Others go further and say that the physics-based models are not fast enough and every key process can be modelled by statistical techniques, so let's throw out the physics (along with the earth scientists and engineers). I do not dispute the incredible power of AI and machine learning, but I hope that a happy compromise of merging human and artificial intelligence into a more power partnership will emerge from all the advances coming today. What that compromise is will be determined by future experimentation, but I am not ready to throw out all the humans just yet.

Another quote from that Economist article shows that we have to make digital literacy a lifelong learning objective. Technology doesn't stop, so humans can't stop. However: "the lifelong learning that exists today mainly benefits high achievers – and is therefore more likely

to exacerbate inequality than diminish it." Achievement of acceptable digital literacy is not a static goal but a moving target. There are some trends that will help us with this critical challenge. Keep learning folks.

Despite trying to deal with a surplus of graduates versus job offers coming off of the recent commodity price downturn, universities in North America and western Europe are trying to prepare students for the Oil and Gas industry's newfound interest in all things digital. In a tough job market, data skills could offer a critical edge. If two people are interviewing for a job in unconventional exploration, and one has analytical skills and experience and the other does not, "guess who gets the job," said Andy Flowers, director of advanced analytics for Marathon Oil.

The tools are not the hard part of this transition. For traditionalists, the challenge is accepting that some aspects of petroleum engineering (and earth science) are better understood statistically using regression analysis

than by the physics-based petroleum engineering they learned in college. The future will be a mix of physics and statistical analysis. Unconventional E&P forced the issue because the behavior of wells in ultra-tight rock often could not be modeled using classical models. But traditional thinking and experience is needed to target the problems likely to yield the biggest payoff, and reject results which are not physically possible.

Corporate training programs can be divided into those seeking to be power users – workers seeking a deep understanding of tools such as Python, a widely used programming language for data analysis – and those who want to be proficient day-to-day users of tools needed to search and process data (using a tool like Spotfire). I am taking for granted that readers of the Foundations journal are well acquainted with the PPDM data management certification program so I won't go into any additional detail in this article. I didn't want to gloss over this important opportunity for working data managers in the industry.

If you are an engineering student looking for employment in the industry you need to add data and programming to your CV, here are a few opportunities to improve your digital literacy:

**Academic**

- University of Southern California is offering a graduate course in Energy Informatics as part of their CiSOFT (Center for Smart Oilfield Technology) MSc degree specialization. (<https://cisoft.usc.edu/education-program/certificate-of-smart-oilfield-technologies/>)
- Colorado School of Mines is offering an undergraduate minor in data analytics in their Petroleum Engineering department, including a course in Petroleum Data Analytics. (<https://catalog.mines.edu/undergraduate/programs/earthscieng/petroleumengineering/#minortext>)
- University of Texas at Austin, petroleum engineering students take courses on analytics. (<https://www.pge.utexas.edu/>)


- Rice University's Wiess School of Natural Sciences is opening up a new Energy Data Management Focus within the M.Sc. in Subsurface Geoscience ([www.profms.rice.edu](http://www.profms.rice.edu))
- The University of Tulsa has an online Master of Energy Business degree designed for working professionals ([www.utulsa.edu/meb](http://www.utulsa.edu/meb))
- The University of Aberdeen is offering a M.Sc. in Petroleum Data Management. (<https://www.abdn.ac.uk/study/postgraduate-taught/degree-programmes/1009/petroleum-data-management/>)
- Robert Gordon University in Aberdeen is offering a degree certificate in Petroleum Data Management. (<https://www.rgu.ac.uk/study/courses/687-gradcert-petroleum-data-management>)
- IFP in France is also offering a degree program in Petroleum Data Management. ([www.ifp-school.com/jcms/r\\_17844/en/specialized-master-s-petroleum-data-management?hlText=Petroleum+Data+Management](http://www.ifp-school.com/jcms/r_17844/en/specialized-master-s-petroleum-data-management?hlText=Petroleum+Data+Management))
- Emerson and Texas A&M announced the establishment of the Emerson Advanced Automation Laboratory to be funded with the company's \$1.5 million donation. The laboratory will provide Texas A&M engineering students a modern, high tech, active learning environment, simulating real-world plant operations found in manufacturing facilities for the oil and gas, refining, life sciences, food and beverage, and other industries. The laboratory will be an integral part of the university's new Zachry Engineering Education Complex, a 525,000-square foot, state-of-the-art facility scheduled to open this fall. ([www.emerson.com/en-us/news/automation/1806-tamu-engineering-donation](http://www.emerson.com/en-us/news/automation/1806-tamu-engineering-donation))

#### Professional Societies

- Professional Petroleum Data Management (PPDM) Association ([www.ppdm.org](http://www.ppdm.org))
- SPE Digital Energy Technology Section (DETS) ([www.spe.org](http://www.spe.org))

- SPE Gulf Coast Section; Data Analytics Study Group. (<https://www.spegcs.org/study-groups/data-analytics-initiative/>)

#### Corporate

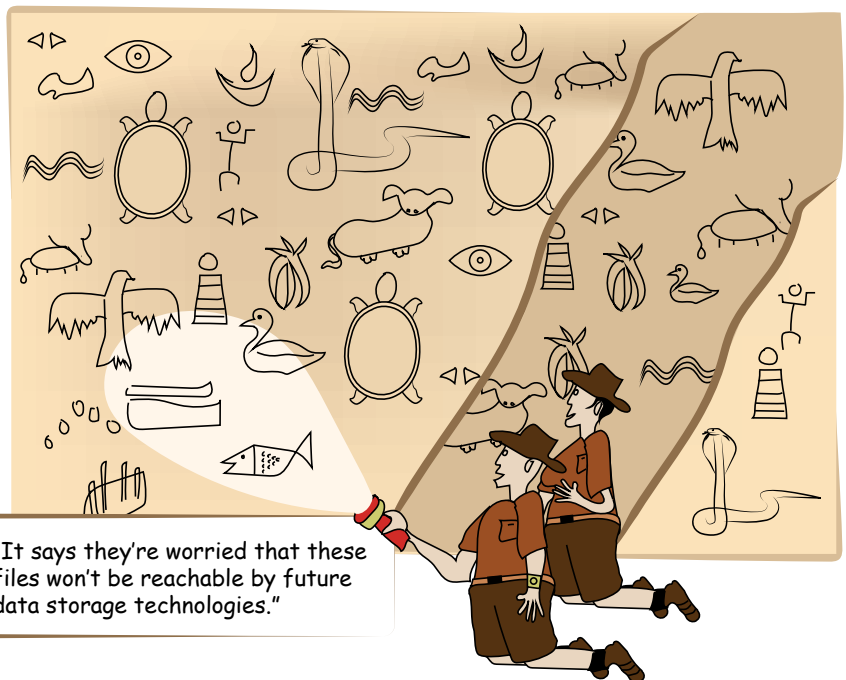
- BP data science bootcamp ([https://www.bp.com/en\\_us/bp-us/media-room/multimedia/videos/well-advisor-technology.html](https://www.bp.com/en_us/bp-us/media-room/multimedia/videos/well-advisor-technology.html))
- ConocoPhillips Citizen Data Scientist program
- Halliburton, "Leading the Way Transforming Talent for E&P 4.0", Big Data and Data Science Bootcamp, Dr. Satyam Priyasarshy & team. 

#### About the Author

Jim retired from Chevron in 2013 after almost 37 years. After retiring, Jim established Reflections Data Consulting LLC to continue his work in the area of data management, standards and analytics for the exploration and production industry.



#### ON THE LIGHTER SIDE



Humour courtesy of Yogi Schulz, Corvelle Consulting

Simplified Data Management.

---

Dynamic Insights.

---

Value-Driven Results.



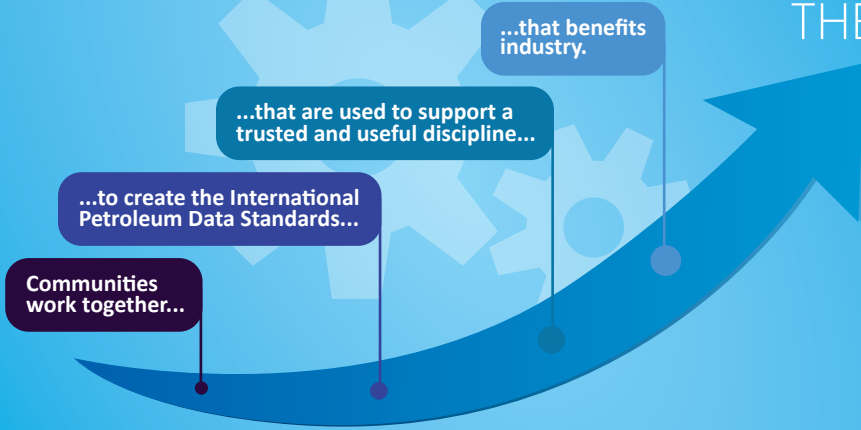
**EnerHub™** is the game-changing oil and gas data management solution built on Stonebridge Consulting's 20+ years' experience in driving operational excellence in oil and gas.

> Learn more at [www.sbti.com/enerhub](http://www.sbti.com/enerhub).



*Business advisory and technology solutions for next-gen oil and gas*

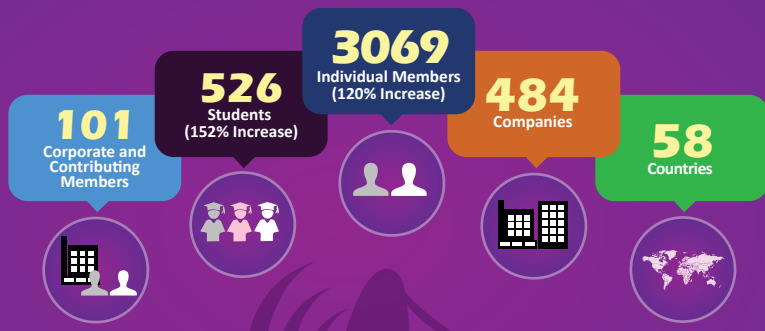
[www.sbti.com](http://www.sbti.com) | [info@sbti.com](mailto:info@sbti.com) | 800.776.9755



COMMUNITY

## MEMBERSHIP

11,000+ Connections



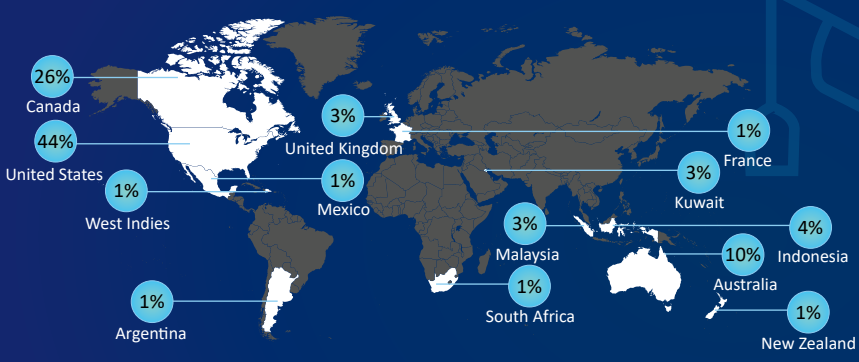
## EVENTS

Number of Attendees	2,034
Number of Sponsors	108
Number of Events	31
Number of Cities	12



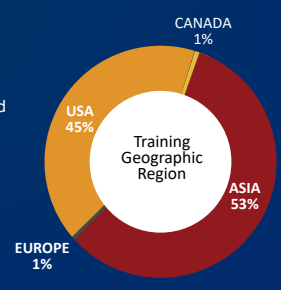
## CERTIFICATION

CPDA Certified by Country



## TRAINING

Number of Seats Last 5 Years	2,698
Number of Seats Current Year	1,240
Number of E-learning/Webinar	632



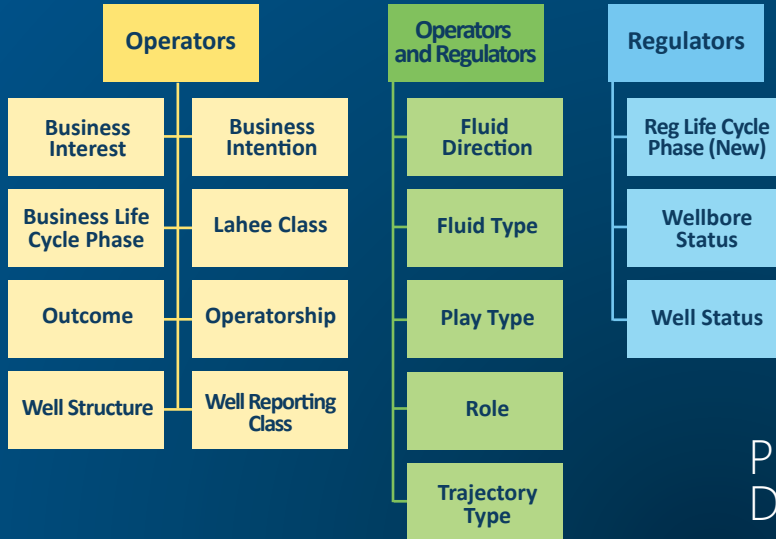
PROFESSIONAL DEVELOPMENT

- Most Popular Classes
- Well Status and Classification
- What is a Well?
- Data Model Overview
- Implementation Head Start
- Data Management and Reference Lists
- Business Life Cycle of the Well Level 1.2

177  
Status Values

16  
Well Status Facets

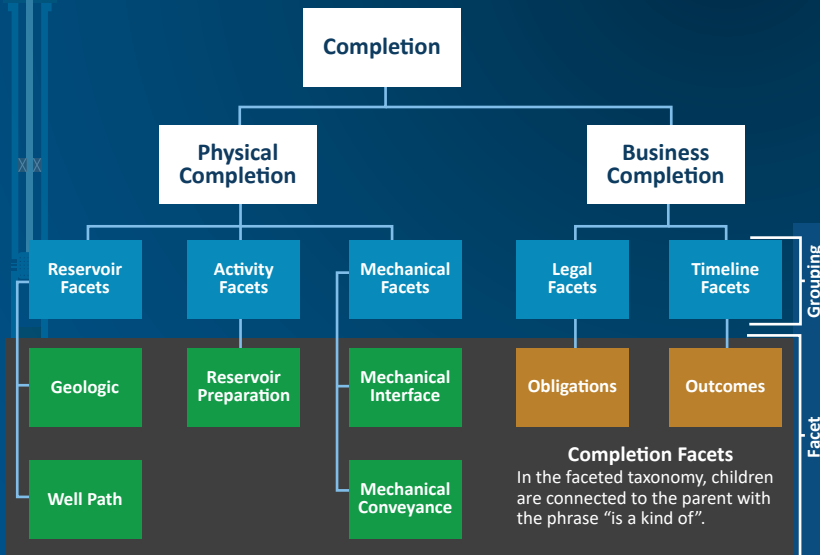
V3  
Out for review



PPDM 3.9  
DATA MODEL

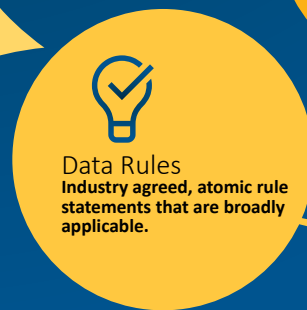


WHAT IS A COMPLETION?

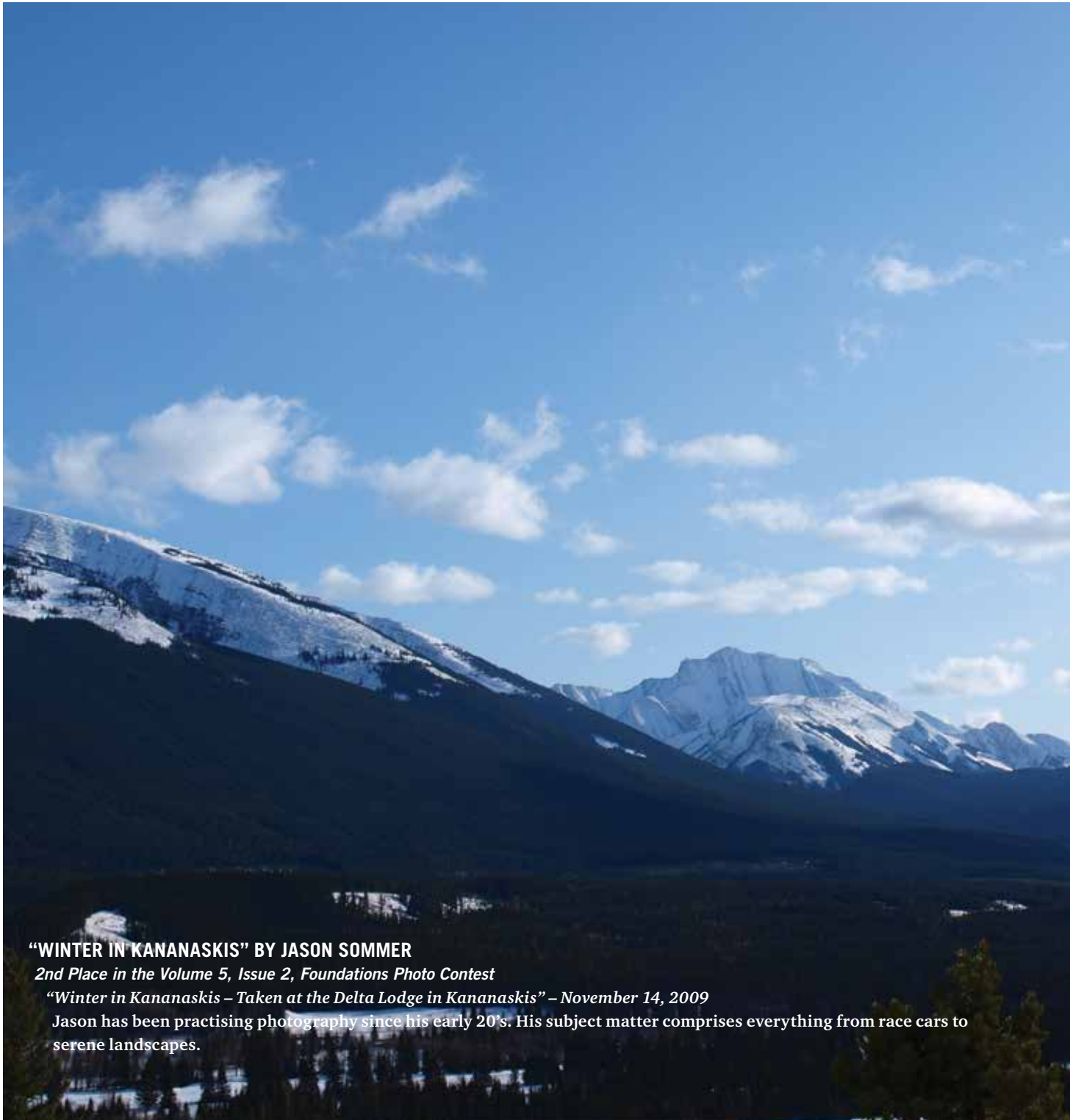


PPDM RULES LIBRARY

3,500  
Rules



## Foundations photo contest



**“WINTER IN KANANASKIS” BY JASON SOMMER**

*2nd Place in the Volume 5, Issue 2, Foundations Photo Contest*

*“Winter in Kananaskis – Taken at the Delta Lodge in Kananaskis” – November 14, 2009*

Jason has been practising photography since his early 20's. His subject matter comprises everything from race cars to serene landscapes.



**On the cover:**



**“OFFSHORE DRILLING IN MALAYSIA”  
BY JIM BOUD**

*1st Place in the Volume 5, Issue 2, Foundations*

*Photo Contest*

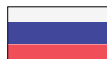
*“Offshore drilling near the Selat Melaka Mosque in the Strait of Malacca, just off the west coast of Malaysia.”  
– October 3, 2012*

*Jim Boud is a Data Architect in the Oil & Gas Industry and doubles as an amateur photographer when on business trips.*

Enter your favourite photos online at [photocontest.ppdm.org](http://photocontest.ppdm.org) for a chance to be featured on the cover of our next issue of *Foundations*!

## ETHICS OF BIG DATA AND OPTIMIZATION TOOLS: Aspects to take into account to measure performance and to set targets in drilling and completion operations

By Ramón Perdomo PhD(s) & Melekhina Marina PhD, Ukhta State Technical University



**B**ig Data is creating a revolution in many areas of our daily life and generating new considerations for our values and behaviors. New technologies are allowing engagement, interaction and participation of a greater number of people. One outcome of the growing presence of Big Data technology is that business operations are changing and increasing the sheer amount of information they generate so fast that the Big Data phenomenon is starting to raise ethical questions,<sup>[3]</sup> including concerns about the handling of information and its purpose.

### THE USE OF BIG DATA

We consider it appropriate to begin by defining some concepts applicable to the ethics of operational optimization. Optimization pursues the best way to achieve a particular objective (or multiple objectives) related to that process under a number of resource or other

constraints. Particularly, the goal of many drilling and completion operations is to reduce cost and deliver predictable results – results that meet the planned cost, schedule and functionality.<sup>[24]</sup>

Risk management and ethics of operational optimization depend on each other.<sup>[8]</sup> Good risk management requires good ethics; and good ethics require good risk management.<sup>[8]</sup> This implies, from a positive perspective:

- First, for an organization to manage its risks well, everyone who represents that organization must practice good ethics.
- Second, for an organization to act ethically, everyone who represents that organization must manage risk.

While Big Data technology offers the ability to connect information and innovate new products and services for both profit and the greater social good, it is, like all technology, ethically neutral. That means it does not come with a built-in perspective on what is right or wrong or what is good or bad in using it. Big Data technology has no value framework. Individuals and corporations, however, do have value systems, and it is only by asking and seeking answers to ethical questions that we can ensure big data is used in a way that aligns with those values. Below is a framework for data ethics:<sup>[3]</sup>

- **Identity:** What is the relationship between user identity (online) that generates the data and the identity of the real person (offline)?
- **Privacy:** Who should control access to data?
- **Ownership:** Who owns data, can rights to it be transferred, and what are the obligations of people who generate and use the data?
- **Reputation:** How can we determine what data is trustworthy? Big data exponentially increases the amount



### SOME USEFUL DEFINITIONS

- **Big Data** is an expression coined to represent an aggregation of datasets that are voluminous, complex, disparate and/or collated at very high frequencies, resulting in substantive analytical difficulties that cannot be addressed by traditional data processing applications and tools (Holdaway, 2014).
- **Model-driven engineering** is a software development methodology which focuses on creating and exploiting domain models. That is, abstract representations of the knowledge and activities that govern a particular application domain, rather than on the computing (or algorithms) concepts (Cordova, 2012).
- A **mathematical model** is a representation of a natural phenomenon or system using variables and mathematical operators to represent components and their interrelationships, which is used to generate knowledge and insights into the system (Eykhoff, 1974).
- An **algorithm** is any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output. An algorithm is thus a sequence of computational steps that transform the input into the output (Cormen, et al. 1990).
- **Process optimization** is the discipline of adjusting a process to optimize some specified set of parameters without violating some constraint. The most common goals are minimizing cost, maximizing throughput, and/or efficiency (Taguchi).
- **Process optimization tool** refers to the use of statistical techniques to identify the optimal solution (Taguchi).

of information and ways we can interact with it. This phenomenon increases the complexity of managing how we are perceived and judged.

- **Algorithmic accountability:** Does transparency and understanding of the purpose of mathematical models exist and how it is being deployed in specific applications and over repeated interactions? Is the decision process clear?

## METHODS USED TO APPROACH A PROBLEM IN ENGINEERING – DATA AND PROCESS OPTIMIZATION TOOLS

The Oil and Gas industry collects large amounts of data from sensors both surface and subsoil. Well drilling and well completion are increasingly complex activities that require continuous OPEX reduction and identify room for improvement opportunities in order to reduce Operational Times (OT), to minimize Non-Productive Time (NPT), and Invisible Lost Time (ILT).

The characteristics of Big Data: Variety, Velocity, and Volume, have put enormous pressure on the organization of databases and data mining. The increasing amount of data pushes fundamental changes in the way data is collected, stored, analyzed and accessed to support intelligence in real time and condensed decision-making cycles. The volume and types of data that are being collected during drilling and completion operations continues to increase. But new data does not result in new performance unless its existence actually changes the way the work is done.<sup>[23]</sup> The manner in which data is used varies largely with the types of decisions that will be made with the information, particularly whether it is intended to maintain current practices or to change the practices.<sup>[23]</sup>

According to Holdaway,<sup>[12]</sup> the three tenets of upstream data are:

- Data management,
- Quantification of uncertainty, and
- Risk assessment.

Among the major upstream operational issues from a safety point of view,

are drilling and completion. Recent developments have allowed the industry to shift and improve its technological base to enhance its safety record. Advances in the industry include the incorporation of modern technologies such as machine learning for fault detection and minimizing human errors.<sup>[7]</sup>

Maidla et al;<sup>[14]</sup> have explained how drilling analysis using big data has been misused and abused. Many people in the Oil and Gas industry are incorrectly utilizing big data to produce correlations that attempt to identify operational “sweet spots”<sup>[14]</sup> without taking into account:

1. The sensors involved and their limitations;
2. The errors in the placement of these sensors;
3. The frequency of the data and how this impacts the analysis;
4. The quality of the data itself;
5. The appropriate filtering of data to ensure apples-to-apples comparisons;
6. The rig state must be known;
7. Understanding of the physics involved.

Pure data mining, without considering the physics involved, limits the full understanding of what is really taking place and can lead to incorrect speculations.<sup>[14]</sup> Numbers can be deceiving, there is a tremendous pressure in the corporate and scientific worlds to convert uncertainty to risk.<sup>[21]</sup>

Simplicity has its virtues, adding complexity to a system can inject more ways the pieces of the system can interact and produce unpredictable outcomes.<sup>[21]</sup> Kleijnen<sup>[15]</sup> states that a mathematical model itself has no morals; it is an abstract, mathematical entity that belongs to the immaterial world. However, such a model reflects an existing or planned system in the real world; the goal of this modeling is to solve a problem in that world (in order to improve society as a whole or one of its groups such as a particular company). Any model—be it mathematical (computerized) or mental (conceptual)—is based on particular simplifying assumptions.<sup>[14]</sup>

Consequently, the model’s results (output) are valid if those assumptions

hold. This consequence leads to the crucial question: What happens when these assumptions do not hold? Often the answer remains unknown, because the modelers do not investigate this question thoroughly; maybe their clients like the answers that the model gives. Yet, an old saying in computer science is “Garbage In, Garbage Out” (GIGO).<sup>[15]</sup>

Digital data strategies are most effective if they support deterministic physics-based decisions rather than statistical or historical practices.<sup>[23]</sup> Model documentation is necessary to enable other researchers (modelers) to reproduce the outcomes of the model. Indeed, reproduction—or its antithesis, falsification—is a basic principle of science.<sup>[15]</sup>

Statistical methods have been used to great advantage to improve product quality or job execution.<sup>[23]</sup> They identify non-compliance or variability. However, a deterministic understanding of how things work, coupled with the appropriated data collection and display, provides the team a type of unique viewpoints that can result in more fundamental changes in how the work is done, which yields greater changes in performance.<sup>[23]</sup>

A strategy to use data must begin with an understanding of the work, the people conducting the work, and the specific characteristics of the organization it is being implemented through. When developing a plan to use digital data, the starting point must be the question of “to what purpose?”<sup>[23]</sup>

## FINAL REMARKS - ETHICS OF BIG DATA AND OPTIMIZATION TOOLS

Ethics are highly conceptual and abstract, but the actions that you take to design and execute big-data innovations in business have very real consequences.<sup>[3]</sup> Damage to your brand and customer relationship, privacy violations, running afoul of emerging legislation, and the possibility of unintentionally damaging reputation are all potential risks.<sup>[3]</sup> Ethical issues in modeling are essential issues for all modelers, because all

modelers are humans and all humans must face moral problems.<sup>[15]</sup>

Learning how to recognize ethical decision points and developing an ability to generate explicit ethical discussions provide organizations with an operational capability that will be increasingly important in the future:<sup>[3]</sup> the ability to demonstrate that business practices honor their values.<sup>[3]</sup> This is, undoubtedly, not the only framework that can help. Other organizational capabilities are required to support those business practices.<sup>[3]</sup>

Finally, we would like to quote the two principles described by Giorgio Gallo,<sup>[19]</sup> as we consider them important aspects to consider when measuring performance and setting targets in drilling and completion operations:

- The first is the “responsibility principle,” proposed in a more general context by the philosopher Hans Jonas, which suggests to take into account in the work not only the point of view of the “client”, but also the point of view of all the “stakeholders.”
- The second, which can be called the sharing and cooperation principle, calls for a more open distribution of the results of our activity, whether they are ideas, algorithms or software. 📄

#### About the Authors

Ramón Perdomo, Sr. Drilling Engineer, student of drilling and well completion technology program at Ukhta Technical State University, his research falls within the modeling and automation of downhole drilling and focuses on how use concrete down-hole data collected (Big Data) and the use of analytical workflows/process to achieving drilling performance in complex wells.

Marina Melekhina is Full Professor at the department of philosophy and teaching methods, Ukhta Technical State University.

Please visit

[www.ppdm.org/foundations](http://www.ppdm.org/foundations)  
for the full article.



By Hazel Welch, Chesapeake Energy

## The Value of a Certified Petroleum Data Analyst (CPDA™) – Worth the Investment

# A

**many companies** look



at ways to invest in technologies to certify and make their data trusted, why not look at the investment they can make on certifying their employees with specific data management roles? The investment companies make on their data should have an equal bearing on the people who support and manage their data.

According to the International Institute of Business Analysis, “professional certification is a designation earned by employees that identifies they have demonstrated a standard level of skills, experience and expertise within their fields. Certifications are generally earned from a professional society that has a certifying body, and are granted based on a combination of education, experience and knowledge, rather than solely by passing an exam.”

There is a lot of value and benefit of having a team of Certified Petroleum Data Analysts (CPDAs) within an organization. It sends a message to others, internal and external to an organization, that they not only value data as an asset, but also that people who manage their data are an asset worth investing in. You don’t just have a data management team, rather a team of professionally Certified Petroleum Data Analysts that have successfully demonstrated high competencies in the exploration and production lifecycle processes, data quality, spatial data knowledge, master data management, data governance, data analysis and data security.

This provides confidence to a more sophisticated working environment and interaction within the team, an exclusive membership to the data analysis “club.”

Studies have shown that companies that support their employee’s professional development and career ladder growth have lower attrition than companies that don’t. Neil Constantine, Managing Director of DataCo Australia, says “we recognize the value of a CPDA™ and see this as a commercial differentiator, allowing DataCo to demonstrate to clients our commitment to data management as a profession and a means to ensure that DataCo staff have the opportunity for continued development and to become well-rounded data managers.”

As more companies undergo digital transformation as the result of Big Data analytics and Artificial Intelligence analysis, having confidence that you have the trusted source of data can be traced to those having the expertise and knowledge in the data. A team that is CPDA certified designates that there is a high-level of expertise and knowledge in the data team. Therefore, your organization can more easily and confidently rely on your team’s ability to meet the petroleum industry’s benchmarks and can be counted on delivering work of a higher and more consistent quality. That’s a peace of mind most companies are willing to invest in. According to Dice.com, the leading website for technology careers, more companies are willing to pay for certifications says Nick Kolakowski a Dice reporter on technology insights and trends. He states that “Some 47 percent of hiring managers say they are



more likely to hire a certified professional than one without a certification.” Several companies have recently included *CPDA Preferred* or *Required* on their data analyst job postings, reinforcing Kolakowski’s statement, and the importance of having certified professionals in their employ.

CPDA teams in an organization can benefit a company as a resource to draw upon when needed. One example would be to have a CPDA attest to a data related litigation case. The certification proves to others that the knowledge in the team or individual handling the data comes with a high degree of professional credibility that only a select few have accomplished, mastering the body of knowledge around petroleum data management. This makes it difficult to challenge competency in the data as certificants are recognized as the subject matter experts for the organization and have taken great responsibility and measures to keep up with, and speak to, industry standards with confidence. A CPDA certification differentiates your



company from the competition.


A CPDA certified team offers a high return on investment (ROI) back to the organization. Upon successfully completing the exam, certificants can provide mentorship for other teams in the organization extrapolating from their experience and expertise to share important knowledge and learnings to the rest of the organization.



The Certified Petroleum Data Analyst exam is administered using Computer-Based Testing (CBT) therefore the exam is completed online in a virtually proctored environment. This method of administration allows PPDM to offer the same exam experience to candidates regardless of their location globally.

The CPDA exam will be offered on four exam dates in 2019, beginning with an exam on **February 27, 2019**. The application deadline for this exam is January 16, 2019.

Thanks to the Petroleum Data Management Certification Committee

(PDMCC) for their contributions to this article. The PDMCC is made up of subject matter experts dedicated to advancing professional development for the data management community through certification programs. 

**About the Author**

Hazel Welch co-authored the CPDA exam as part of PPDM’s Petroleum Education Task Force and currently serves on the board of PPDM’s Governance Committee. She has been employed with Chesapeake Energy since 2002 in various roles and currently holds the position of Operations Data Advisor on their Data Strategy and Governance Team.

For more information about the Certified Petroleum Data Analyst (CPDA) please visit [www.ppdm.org/certification](http://www.ppdm.org/certification) or contact the registrar ([certification@ppdm.org](mailto:certification@ppdm.org))

Upcoming CPDA Exams				
<b>Exam Date</b>	February 27, 2019	May 22, 2019	August 28, 2019	November 13, 2019
<b>Application Deadline</b>	January 16, 2019	April 10, 2019	July 17, 2019	October 2, 2019

# Two Database Giants

By Amii Bean, EnerVest Operating

**W**hen referring to databases and database technology, there are two main classifications, SQL and NoSQL, known as relational and non-relational databases, or even widely used as structured and unstructured data collections. “We call them SQL and NoSQL, referring to whether or not they’re written solely in structured query language (SQL)” (Wodehouse 2018). Aside from the language(s) used to interface with the data, relational and non-relational databases differentiate in many ways. The schema or data models are on opposite sides of the spectrum; essentially, SQL has one and NoSQL has none. This has led to the two database communities having this continuous battle on which is the better, superior or even proper DBMS (database management system). Why you ask? The more data we collect, store, process, analyze, and visualize, the larger and exponentially more complicated the database will become. Whether you’re a Database Analyst, an Enterprise Architect, a Data Scientist, or a Data Manager, you want to make certain you have the right tool in place to manage your data properly. Data management is becoming more and more important in a vast number of industries, and with the ever-expanding amount of technology available, growing questions arise on how to manage data. How do you know what the best model is to use? How do you know when to use that model? And even, is there a simple equation on which to use and when? Understanding the significant differences and similarities

between SQL and NoSQL will drive the confident decision you’re looking to make. Let’s take a look at the two.

### SQL, THE ORIGINAL

To understand what SQL does, you need to understand the history of it. The relational data model was born in the 1970s by Ted Codd, who wrote a paper in order to start the conversation on and even try to define how data is related. SQL (Structured Query Language) soon emerged, at IBM no less, developed by a couple of American scientists during that same timeframe after learning about the relational model from Ted’s research. The initial version was originally called SEQUEL, or structured English query language, and was designed to retrieve data and manipulate data within the original database management system of IBM resembling that of a relational database. However, that original version relied on complex mathematical logic and equations, and researchers and scientists really “wanted a query language that was as easy to read as English, and...would also encompass database administration and manipulation” (Kulkarni 2017). IBM soon thereafter released its first commercially available SQL product, and now SQL has become the predominant language for interacting with a database. SQL’s foundation works through a series of queries and transactions for data structured in rows and titled columns usually consisting of text, numbers, and dates. Each column has a specified data type as this is used to understand



what type of data is expected in a given column. The defined data types are also the way to identify how SQL will interact with the stored data, i.e., queries and transactions. A query is a single statement that uses one command to either read or modify data. Transactions are performed through a continuous sequence of statements against the database in order to maintain consistency in the data and durability of the database. They “ensure only one change can be written to a data field at a time, so [no] conflicting transactions made” (Vaas 2016). Ensuring ACID (Atomicity, Consistency, Isolation, Durability) compliance is a strong advantage of SQL databases. What this means is, it reduces anomalies and protects data and database integrity. This compliance is used in the transactions when querying and modifying the data, so either the entire transaction is completed or none at all, thus using the famous ‘rollback’ and/or ‘commit’ method. The difference between a query and a transaction can be summed in a single statement. A query is like a single page of a book, while a transaction is like an entire chapter. SQL boasts structure as an advantage; so the data has to be arranged in a very organized and normalized way. Having “a well-designed schema minimizes data redundancy and prevents tables from becoming out-of-sync” (Wodehouse 2018). Because of the way the data is sorted and stored in a patterned, structured format, SQL makes it easy to search and change pieces the data collection and makes it easier for big data applications



to digest the data, or so it seems.

**WHAT IS THIS NOSQL THING?**

So SQL looks at structured data in a relational database, but what about everything else? SQL and NoSQL do the same thing; they store data. Yet, there is more than just tables and rows grouped together in a string format for data storage. There are audio and video, email and documents, web and even social media data that is all unstructured. The internet indeed invaded our everyday lives and that of the technology industry including the data management world in countless ways. From that disruption NoSQL (or Not Only SQL) appeared, and the threat of SQLs existence began. “As an alternative to tables, NoSQL databases use a document-oriented approach to database management. This means that non-structured data (such as [said] text[s], images, social media data, videos, etc...) can be stored in a single document” (NewGenApps 2017). These unstructured databases do not have a schema, or model to follow, and are more dynamic in nature, so it will be easy to write data to the database and locate files within the database. This is definitely a more intuitive way to handle data, but this will require more processing and more storage from a database. However, because the data isn’t stored in fields or categorized like that of a relational database, “NoSQL databases are often associated with big data tasks [rather than data digestion], handling large volumes of data in various forms” (Vaas 2016). NoSQL is still uncommon to most

and is not a defined, concrete concept. However, its features and variants in functionality help to solve some of the key issues of SQL databases. Because of its flexibility, NoSQL is massive in scalability so you can increase the number of servers with the escalation in traffic. NoSQL databases have proven to handle an extreme amount of data and the variation of data types. Yet with the lack of orderly internal structure, “users can [only] run simple content searches across textual unstructured data...and the enterprise gets little value from potentially valuable data sources...” (Taylor 2018). However, there is so much more unstructured data than structured, and the only instability

of NoSQL revolves around the uncertainty of data analytics tools and the many ways to write the query language to interface with the data. “Unstructured data makes up 80 percent, and more, of enterprise data, and... without the [developed] tools to analyze this massive data, organizations are leaving vast amounts of valuable data on the business intelligence table” (Taylor 2018). But is this a replacement for SQL, or is it merely an alternative?

**SUMMARY**

When evaluating a project for data management, think about the larger picture at hand. Ensure you are going to use the right tool for the job. NoSQL and SQL were designed to solve different problems, so forcing one way or the other based on preference or because one is newer, shinier, and more powerful could yield potential for project failure. Do your homework, understand what the project is trying to accomplish and act accordingly. **■**

**About the Author**

Amii Bean, EnerVest Operating. Amii has managed data for several E&P companies over the last 14 years, with her latest role as the data administration liaison between engineering and information technology.

**AN ODE TO DATA MANAGERS**

*With apologies to William Shakespeare*  
 By Jess Kozman, Woodside Energy

When naughts and zeroes grow to fill our drives  
 And users quake as each new wave arrives  
 That’s the time when cooler heads prevail  
 And only data lives to tell the tale.

For years untold now we have talked and planned  
 While those we knew retired or were canned.  
 But now the fruits of all those toils hang ripe  
 And even CEO’s repeat our hype!

So gather round my fellow data nerds  
 And mark this well and heed our every word.  
 For while we may not be so wise or sage  
 For us this truly is, a golden age.



## Thanks to our Volunteers

### MAY 2018

#### Sean Udell

May's Volunteer of the Month is Sean Udell. For many years, Sean has been a valued part of the PPDM community in a variety of roles, including several PPDM work groups (*What is a Well, Well Status and Classification, Global Framework for Well Identification, US Well Number Standard, and the Canadian Well Identification System Standard*). Sean also represented geoLOGIC systems ltd. on the Sponsor's Steering Committee for the overall *Well Identification* set of projects. Sean currently sits on the Regulatory Data Standards Committee, and recently presented an update on the work of the committee at the 2018 Houston Expo. "Sean is always there for our team, whether it's presenting at a conference, hosting a training, or even helping us with preparing printed materials. His knowledge and experience has been invaluable, and we are delighted that he's our Volunteer of the Month," said Ingrid Kristel, Senior Project Manager.

Sean has more than 25 years of experience in the oil and gas industry, all with geoLOGIC systems ltd. Starting as Canadian Sales Manager in 1991, in 1995 he shifted to product design, product documentation, and product support focus as the Operations Manager. In 1997, he became Vice-President, Operations and Technology, where his responsibilities included overseeing product design for geoSCOUT, the geoLOGIC Data Center (gDC), and many web-based products and services. Sean recently attained his Certified Petroleum Data Analyst (CPDA™) credential.

### JUNE 2018

#### Terence Broughton

Congratulations to Terence Broughton, PPDM's June 2018 Volunteer of the Month.



Terry is a Certified Petroleum Data Analyst (CPDA) with more than 20 years' experience in petroleum data management as an independent contractor and president of TB Data Solutions Inc. After earning his CPDA, Terry joined the Petroleum Data Management Certification Committee (PDMCC).

"Terry has been a fantastic addition to the PDMCC, bringing energy and enthusiasm to the group. Terry has been helping to improve the candidate experience so that we can continue to offer a great experience to our candidates and the prospective CPDAs. We truly value his experience and wisdom," said Ingrid Kristel, Senior Project Manager.

### AUGUST 2018


#### Daniel Perna

Daniel Perna is PPDM's August 2018 Volunteer of the Month. Daniel is a data management professional at Devon Energy, where he has worked for 10

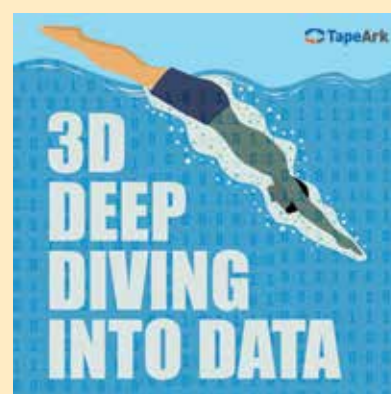


years. As part of the Geoscience and Well Data Management Team, he does business analysis, data analysis, coordination, and training for stakeholders across disciplines. He also applies his subject matter expertise to standards, best practices, solutions development, and the improvement of data quality and integrations. Daniel was also one of PPDM's first Certified Petroleum Data Analysts (CPDA).

"Daniel joined the Oklahoma City Leadership Team in 2016 and has since taken on the Chair's role with ease," said Pam Kosciński, PPDM's USA representative. "Under Daniel and

the Leadership Team's hard work, the Oklahoma City community has grown, and our Workshop numbers have nearly doubled. Additionally, Daniel has spent countless hours putting together materials to help CPDA candidates and mentoring them. Daniel is a real asset to PPDM." 

### PERTH WORKSHOP PODCAST



The record breaking PPDM Perth Data Management Workshop & Field Trip was also the first PPDM event to have a Podcast produced.

The podcast, produced by Tape Ark in Magic Studios Perth, features an in depth interview with Trudy Curtis the CEO of PPDM, as well as back-stage interviews with conference presenters on a wide range of topics related to the work of PPDM, the establishment of standards, training, certification, and general updates from global leaders in data management. The Podcast entitled "Tape Ark 3D – Deep Diving Into Data" was generously donated to PPDM, and will be available at <https://www.tapeark.com/tapearkpodcast/> after September 1, 2018. Tape Ark welcomes feedback on the production and commentary on the interest to continue podcast of a similar nature.

# Hands On With The PPDM Association Board of Directors



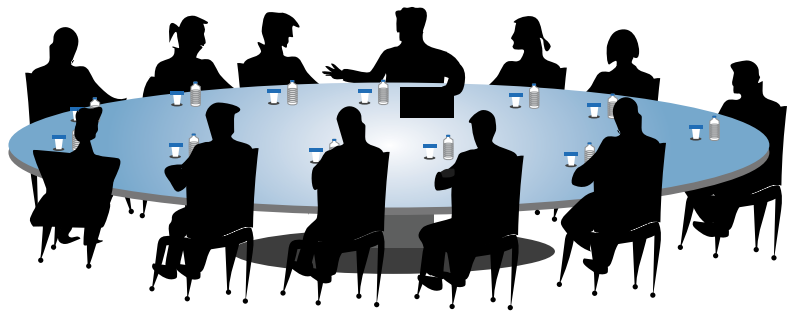
## To SQL or NoSQL: That is the Question

By Amii Bean, EnerVest Operating

**S**QL (Structured Query Language) has been around since the 1970s and the dawn of relation database management systems. Over the next few decades, SQL proved to be immensely popular in the data industry. Then the Internet happened and many data types were needing to be stored in something other than tables and rows. So appeared NoSQL and its massive storage capabilities. This article explores the similarities/differences and advantages/disadvantages between the two main designations of databases and why an enterprise would choose one over the other. This article also briefly touches on where technology may take the data industry in the near and broad future.

### WELL-KNOWN ADVANTAGES

Both classifications of databases have their place in the data management world, and both hold a great deal of value. SQL is more widely adopted and used, and the tools have more support given the time it has been on the market. If you are using data that is consistent or using data that must be consistent without any room for error, utilize SQL and a relational database. If your business is not experiencing the massive growth that would require increasing the capacity of the servers the database sits in, utilizing a SQL database is the way to go as there wouldn't be a need to bring in a massive system that supports all types of data. However, if you want to store large amounts of data that have little to no structure, use a NoSQL database. These types of databases do not set limits on



the types of data you store and allows you to continue to add different types of data within the same database. NoSQL also doesn't enforce a definition on the data type before it is stored. Therefore, if you need to set up a system quickly with the agility of scaling frequently, SQL will slow you down, and NoSQL would be the better option as it requires zero prep time and has the flexibility of increasing its capacity by increasing "the pool of resources to reduce the load" (Issac 2014). Additionally, because their schemas allow changes without needing to stop the database, they are much more adaptable to the project. "At this point it's very important to insist on the fact that even though it may look like a good time to move over to a NoSQL database, we must keep in mind, before taking that decision, [of] the characteristics of our database rather [than the idea that it] needs a NoSQL or a relational database" (Javier 2016). Some projects are better suited for NoSQL; some need the structure of SQL, not one database classification will solve your enterprise's problems or needs.

### KEY DISADVANTAGES

SQL is proven to be the go-to database mainly for the popularity and advantage of the known aspects of it. Yet, there are things that prevent it from being the all-encompassing data management system. SQL has an issue with performance due to the atomicity of transactions and operations in the database. What this means is, the more you query and modify the data, the slower the database will perform at that time. In order to 'fix' this performance issue, you need to increase the horsepower of the hardware the database sits on. To do that, you'll need to make the server more robust, but you can't do that without taking the database down (given we are talking only about a smaller enterprise's set up). That entire process would end up costing a company a substantial amount of money, and that isn't feasible in most operations without having an alternative option ready or a planned budget. So, although a relational database has excellent organization in the schema and an immense amount of support and understanding on the structure and language, the lack of


flexibility restricts what an enterprise can offer to the data management end user. There seem to be more advantages to using a NoSQL database based on what was just mentioned, but there are disadvantages as well. NoSQL uses its own query language called Unstructured Query Language (UnQL), and the syntax (code arrangement) varies from database to database. Consequently, the NoSQL community doesn't have strong support for broadly understanding how to pull data from the database. Because NoSQL is better with tasks, it can't handle large complex queries; it just doesn't have the standard interfaces to perform in this manner, and thus are not as powerful as the SQL query language. SQL is built for sizeable transactional type applications, and while NoSQL can handle transactions, "it is still not comparable and stable enough in high load and for complex transactional applications" (Issac 2014). Even though the querying capability isn't as strong as it should be in a NoSQL database, and with the advancement of API (Application Programming Interface) in the data management world, does SQL have a fighting chance against the massive power of NoSQL?

### WHAT DOES THE FUTURE REALLY LOOK LIKE?

SQL's downside in the cost to upgrade and shortcomings of inflexibility make it appear to be the inferior database, but the ability to query data allows it to keep a fighting chance. NoSQL lacks the traditional features within a SQL database that developers rely on such as the query language. With NoSQL, developers have to write complex code and aggregate data across the system to retrieve the data in a readable manner, and that isn't conducive to daily operations of a given enterprise. Yes, NoSQL methods are useful in "simple [data] retrieval scenarios, SQL has provided significant additional value in expressing more complex data access patterns and pushing computation to the data" (Kulkarni 2017). We live in a world and an era where data is becoming the

	NOSQL	SQL
MODEL	NON-RELATIONAL STORES DATA IN JSON DOCUMENTS, KEY/VALUE PAIRS, WIDE COLUMN STORES, OR GRAPHS	RELATIONAL STORES DATA IN A TABLE
DATA	OFFERS FLEXIBILITY AS NOT EVERY RECORD NEEDS TO STORE THE SAME PROPERTIES NEW PROPERTIES CAN BE ADDED ON THE FLY RELATIONSHIPS ARE OFTEN CAPTURED BY DEMORMALIZING DATA AND PRESENTING ALL DATA FOR AN OBJECT IN A SINGLE RECORD GOOD FOR SEMI-STRUCTURED, COMPLEX, OR NESTED DATA	GREAT FOR SOLUTIONS WHERE EVERY RECORD HAS THE SAME PROPERTIES ADDING A NEW PROPERTY MAY REQUIRE ALTERING SCHEMAS OR BACKFILLING DATA RELATIONSHIPS ARE OFTEN CAPTURED IN NORMALIZED MODEL USING JOINS TO RESOLVE REFERENCES ACROSS TABLES GOOD FOR STRUCTURED DATA
SCHEMA	DYNAMIC OR FLEXIBLE SCHEMAS DATABASE IS SCHEMA-AGNOSTIC AND THE SCHEMA IS DICTATED BY THE APPLICATION. THIS ALLOWS FOR AGILITY AND HIGHLY ITERATIVE DEVELOPMENT	STRICT SCHEMA SCHEMA MUST BE MAINTAINED AND KEPT IN SYNC BETWEEN APPLICATION AND DATABASE
TRANSACTIONS	ACID TRANSACTION SUPPORT VARIES PER SOLUTION	SUPPORTS ACID TRANSACTIONS
CONSISTENCY & AVAILABILITY	EVENTUAL TO STRONG CONSISTENCY SUPPORTED, DEPENDING ON SOLUTION CONSISTENCY, AVAILABILITY AND PERFORMANCE CAN BE TRADED TO MEET THE NEEDS OF THE APPLICATION (CAP THEOREM)	STRONG CONSISTENCY ENFORCED CONSISTENCY IS PRIORITIZED OVER AVAILABILITY AND PERFORMANCE
PERFORMANCE	PERFORMANCE CAN BE MAXIMIZED BY REDUCING CONSISTENCY, IF NEEDED ALL INFORMATION ABOUT AN ENTITY IS TYPICALLY IN A SINGLE RECORD, SO AN UPDATE CAN HAPPEN IN ONE OPERATIONS	INSERT AND UPDATE PERFORMANCE IS DEPENDENT UPON HOW FAST WRITE IS COMMITTED, AS STRONG CONSISTENCY IS ENFORCED. PERFORMANCE CAN BE MAXIMIZED BY USING SCALING UP AVAILABLE RESOURCES AND USING GIN-MEMORY STRUCTURES. INFORMATION ABOUT AN ENTITY MAY BE SPREAD ACROSS MANY TABLES OR ROWS REQUIRING MANY JOINS TO COMPLETE AN UPDATE OR QUERY
SCALE	SCALING IS TYPICALLY ACHIEVED HORIZONTALLY WITH DATA PARTITIONED TO SPAN SERVERS	SCALING IS TYPICALLY ACHIEVED VERTICALLY WITH MORE SERVER RESOURCES

world's most valuable asset, and as a result "we have seen a Cambrian explosion of specialized databases, data processing tools, etc." (Kulkarni 2017). SQL is by no means perfect, but it is essentially a universal interface. It is the language that most of us in the data management community know and understand well. And while there is further development of NoSQL and other natural language interfaces in the works, "what will those systems then connect to? SQL" (Kulkarni 2017). SQL will be around for a while and ultimately win the war, "not just because writing glue code to kludge together NoSQL tools is annoying..., not just because retraining workforces to learn a myriad of new languages is hard..., not just because standards can be a good thing. But...because the world is filled with data" (Kulkarni 2017), and data is all around us. As we advance in the data management community and our software and hardware systems become smart enough to help us, data itself will continue to grow, as will the importance of data even more so. And as we collect data more and more, the complexity to be able to store, gather, process, analyze and visualize data will

increase in volumes and thus utilizing a combination of different databases may very well be the answer. The key is to use the technologies where appropriate, SQL for consistency and standards and NoSQL for availability and scalability, for real success. That is what data management really looks like in our future. 

#### About the Author

Amii Bean, EnerVest Operating. Amii has managed data for several E&P companies over the last 14 years, with her latest role as the data administration liaison between engineering and information technology.



## Petroleum Data Management As A Professional Discipline: Creating The PPDM's Professional Development Catalogue

By Tracy Heim, Professional Development Committee

Professional Development (PD), the learning activities that an organization provides to maintain and improve skillsets among its workforce, is essential for establishing and sustaining a high standard of excellence within the fields of corporate information and data management. PD encompasses the educational requirements of individual data managers, but also the corporate hierarchy that provides data, data services, and related human resources through a network of data managers, supervisors and other related administrators. Data management professionals within the petroleum industry are often challenged to find the very opportunities that allow them to increase their knowledge base and further advance in their chosen career paths. In response, PD has become an integral component of PPDM's current mandate for elevating the role of Petroleum Data Management as a professional discipline.

In late 2015, PPDM formed the Professional Development Committee (PDC), as a standing committee focusing on initiatives for: establishing the value of professional development, establishing a centralized repository of educational and

professional development opportunities, creating standards and competencies for data management job roles, developing career path recommendations, and harmonizing compensation.

Towards achieving these goals, a sub-committee was struck in 2016 to meet the primary goal of creating a centralized online repository: the Professional Development (PD) Catalogue. Set to debut this fall on PPDM's website ([www.ppdm.org](http://www.ppdm.org)), the PD Catalogue is designed to ensure that data management and data management-related training courses are more readily, and effectively, available to data managers and their peers. As the industry's not-for-profit data management society, PPDM is working to raise awareness about professional development options for our members. Although this is not an endorsement of the course content, the PD Catalogue Sub-Committee will review all training provider applications to ensure submissions meet basic criteria.

### WHAT WE OFFER

The PD Catalogue represents professional development opportunities such as:

- 1. Course Offerings (Classroom or Web-Based)** – designated as petroleum data management-specific or petroleum data management-relevant classroom or web-based learning opportunity.
- 2. Field Trips** – an informal outing of data management professionals that utilize observation for educational purposes (vendor-neutral content). A field trip presents relevant material related to petroleum data management subject matter and includes information such as clearly defined learning outcomes, field trip description, topics covered, unique requirements, and target audience.
- 3. Seminars** – a more active training session, typically targeting more experienced or higher-level students. The instructor will engage students in a more active discussion related to the subject matter being presented.
- 4. Webinars** – providing petroleum data

management-specific or petroleum data management-relevant training in a, vendor-neutral, web-based live or pre-recorded seminar format conducted over the Internet.

- 5. Workshops** - an active learning session in which participants come together to learn new skills. These skills are related to specific aspects of work in petroleum data management and include, within a workshop syllabus, a workshop title, clearly defined learning outcomes, workshop description, topics covered, unique requirements, target audience, and instructor biographical information; void of any software-specific content, promotion or sales.
- 6. Other offerings** - The PD Catalogue is designed to be adaptive as course delivery methods evolve. The Massive Open Online Course (MOOC), for example, is an online course that encourages unlimited participation through web open access. These options emerged from the open education resources movement and are typically offered free of charge. One major outcome from MOOCs is the networking opportunity found throughout the course.

### OUR TARGET STAKEHOLDERS

#### Petroleum Data Managers

As data management becomes increasingly visible in the petroleum sector, it is often difficult for current and aspiring data managers to source professional development opportunities. The PD Catalogue will provide a central location from which data managers can search for professional development opportunities and quickly register directly at training provider websites. Data managers will be able to view available courses, locations, and even instructor biographies.

#### Supervisors and Human Resources staff

Historically, supervisors and HR staff have been challenged to find relevant professional development opportunities for data management staff. The PD Catalogue will provide quick access to these data management professional

development opportunities, so supervisors and HR administrators can more readily add more relevant tools to career path development plans and promote these offerings to data management teams.

Our repository will be found at [www.ppdm.org/training](http://www.ppdm.org/training).

### Training Providers

With a well-defined framework in place for training, learning institutions and related providers will be able to submit petroleum data management and data management-related course offerings to an even broader audience of data managers and aspiring data managers.

For more information on submitting your course offering to PPDM, please visit our website. Starting in the Fall of 2018, applications can be submitted for inclusion in the repository.

### ABOUT THE SUB-COMMITTEE

The PD Catalogue Sub-Committee is comprised of the following volunteers:

Tracy Heim, Data Analyst with Alberta Energy Regulator in Calgary.

Cindy Cummings, Exploration and Physical Data Coordinator with Repsol in Houston.

Siti Zubaidah (Zubai) Abu Bakar, CPDA, Exploration Database Specialist with Repsol in Kuala Lumpur.

Mark Craig, former Chief Information Officer at BP Canada, now Chief Information Officer at Maurice Law Barristers and Solicitors in Calgary.


Margaret Barron, Chief, Professional Development of PPDM, provides support, guidance and leadership to the committee.

### WE WELCOME YOUR SUPPORT

PPDM continuously encourages individuals who share our passion for promoting Petroleum Data Management as a professional discipline to serve on

our committees. If you are interested in participating in the Professional Development Catalogue sub-committee or other PDC sub-committees, please contact [volunteer@ppdm.org](mailto:volunteer@ppdm.org).

### FOR MORE INFORMATION

If you are a petroleum data manager exploring new professional development opportunities or a Training Provider interested in submitting course outlines, please visit our website [www.ppdm.org/training](http://www.ppdm.org/training) or contact [training@ppdm.org](mailto:training@ppdm.org) for further information. 

### About the Author

Tracy Heim is a Data Analyst with the Alberta Energy Regulator in Calgary. She has been a member of the Professional Development Committee since its inception in 2015, and is a data management professional with more than 10 years' experience in the petroleum industry.



The image displays several overlapping screenshots of the Professional Petroleum Data Management (PPDM) website. The primary screenshot shows the course page for 'PDM-003: Well Status and Classification', detailing the instructor (Tracy Heim), course options (Classroom, Private), and a comprehensive description of the course content, including foundational concepts and learning outcomes. Other screenshots show a 'Training Courses' page with a grid of course thumbnails and a 'Training Providers?' page. A large megaphone icon is positioned on the right side of the collage, pointing towards the text 'PPDM's Professional Development Catalogue'.

## PPDM's Professional Development Catalogue

One Person  
Can Do So Much

**Together**

The word "Together" is written in a large, bold, blue sans-serif font. Below each letter, a hand is shown holding it up from underneath. There are eight hands in total, one for each letter. The hands are light-skinned and are positioned as if they are supporting the word. A thick blue horizontal line runs across the bottom of the word.

**we can do so  
much more**

**Now It's YOUR Turn**

The PPDM Association needs volunteers like YOU to continue to identify professional development opportunities, grow certification programs, advance data management standards, and more.

We have immediate opportunities for volunteers with a variety of skill sets and experience in both our Professional Development Committee and Petroleum Data Management Certification Committees.

**Apply Now**

Contact [volunteer@ppdm.org](mailto:volunteer@ppdm.org) to find out how you can make an impact.

# Six Data Testing Practices to Increase Business Intelligence ROI



By Charity Queret, Stonebridge Consulting

**W**hat is business intelligence data testing? Whereas business intelligence (BI) is the process of using technology to analyze data for the presentation of BI deliverables (e.g., graphs, charts, reports, spreadsheets, etc.), BI data testing is the process of validating data accuracy, format and performance of reports, subject areas and security. A robust, disciplined data testing regimen ensures that BI initiatives, whether large or small, are successful.

BI is not new to the Oil and Gas industry. Today, Oil and Gas organizations of all sizes connect, integrate, interact with, communicate, and report business data. Despite this fact, there seems to be a persistent distrust of BI data and reports among business users. A recent TDWI survey on BI's business value found that a large majority of business and IT executives do not trust current investments in data analytics and do not believe their organizations are getting the insights needed to drive business decisions.

That's a startling data point for BI developers like me. One of the key factors for the success of BI projects is the level of trust for the data shown in BI deliverables. Lack of trust in the data reduces user adoption and often results in BI project failures or reduced ROI.

While data testing is a component of every BI project plan, the practical

implementation of data testing has posed problems ad nauseam. Data testing is typically an "activity" at the end of a BI project or before a helpdesk ticket can be closed. Typically, it consists of a BI developer testing one data scenario, one date period, one format option or security for one data field and then passing the end deliverable on to the business user for their review.

Oil and Gas organizations, like their counterparts in other industry segments, do not always know how to test BI deliverables, what to test or when their test is complete. These poor testing practices are manifested in recurring issues from poor data integrity, formatting, performance and security issues – leading to increased development cost due to late detection of data errors.

The slightest data integrity error can have catastrophic impacts that ripple throughout an Oil and Gas organization. Inaccurate data quality leads to sub-optimal decisions and even to regulatory risks. Discovering issues early in the process reduces the time, effort and cost of correcting them later in the process.

## THE EVOLUTION OF DATA TESTING

Regardless of industry, most BI testing energy is not spent testing the actual data; rather it is spent testing the software functionality that uses or creates the data. Years ago, the software solution itself would be the only source using the data, so if the

application was correct, so was the data. As BI has evolved, however, it is now the norm to have multiple software applications using, storing, creating and manipulating data. This increases the complexity of BI testing, as developers must now test not only the validity of the data but also the inter-connectivity of disparate software using and changing data with cross-interactions.

There are three key phases to a data-centric testing model that validates the entire flow, from data retrieval at the source system, to the end BI report used by the organization.

1. **Data Warehouse/ETL Testing**
  - Basic validation testing of the extract and verification of the data transformation:
    - a. Source and destination data types should match.
    - b. Primary keys, foreign keys and default values should remain intact.
    - c. Data loss and data anomalies between the source and destination should be addressed.
2. **Load Performance Testing** - Key indicators used to test the data storage system:
  - a. Performance
  - b. Scalability
  - c. Error logging
  - d. Exception handling
  - e. Audit trails

3. **Report Testing** - Testing of the last layer of the data flow through validating usability of the report:

- a. Report layouts per mockup
- b. Data validation per business requirements
- c. Filters and prompts
- d. Drill downs and drill throughs
- e. Aggregations
- f. Formats

Payoffs from quality BI data testing enable us to make changes with confidence and release a BI deliverable that assures business users that the data is accurate and that the core expectations of the business community have been met.

**SIX DATA TESTING PRACTICES**

Here are six data testing practices that will ensure testing is not just an afterthought but an integral part of the BI process and user acceptance.

- 1) **Remember the Big Picture** – The goal of BI data testing is to create a trusted and high quality BI deliverable. Testing is not about checking things off a list but rather about taking steps to ensure that a better end product is delivered. Keep the business community in mind, always considering the BI deliverable from their perspective.
- 2) **Allocate Time to Plan** - To keep an eye on the big picture and test with the end goal of a trusted deliverable in mind, you need to have a complete understanding of the initial request and the business community’s expectations for the deliverable. You must take time to identify what the deliverable will be used for, known issues regarding the data, what problem the deliverable will be solving, who will use the deliverable and how the deliverable must look and feel. It is essential to give yourself the time to ask questions to make sure you know everything you need to know about the expectations of the business for the deliverable.
- 3) **Create a Test Plan** - A well-written test plan communicates an



understanding of what will be tested and what is required by the business community to close the request. Specifically, it details how data will be validated, what standards have been defined for format and performance of reports, subject areas to be tested, security measures required, who will validate, and when it will be validated. It is a clear picture of what is required for business user acceptance.

- 4) **Practice Good Testing Documentation** - Documentation of what has been tested and where you are in the test process saves you time and money. Properly managed testing documentation should be able to quickly track where you are in the testing process and what issues you have incurred. It promotes quality assurance by providing actionable insight on risk, progress, results and more.
- 5) **Recognize Silo Mentality** - Successful BI data testing is a “dependent” process and requires collaboration between IT and business users. It is marginal at best to have only the BI team perform data testing. The BI team can detect data issues,

but the business community knows the data best, and because they are the actual consumers of the data, their participation is integral to successfully validating the data within the system.

- 6) **Trust Your Gut** - Not everything can be detailed in a test plan. Sometimes you must use your gut and look for data that doesn’t belong or doesn’t look right. However, be vigilant not to go down a rabbit hole and instead stay focused on the end game.

Data testing is more than an activity.

It is a process and a mindset. It first starts with recognizing that the end game is a quality BI deliverable. If we treat testing as a process that supports the entire BI development process, project or helpdesk ticket, instead of a single activity or phase, we will ensure our data is dependable and robust, renew the trust of business users in BI data and ultimately increase business intelligence ROI. 📌

**About the Author**

Charity Queret is a senior consultant at Stonebridge Consulting. Charity has over 20 years of experience in designing and developing end-to-end business intelligence and data warehousing solutions.

**THE RULES LIBRARY GROWS AGAIN**

Recently, 474 rules were added to the Rules Library. They deal with data quality in wellbore directional surveys, core sample analysis, pipelines, reserves and production reporting. Many of the rules are about ensuring compliance to commonly-used standards or guidance for reports.

These new rules have “In Review” status at [rules.ppdm.org](http://rules.ppdm.org). They are available for all members, and guests who create an account. The review stage is the chance for anyone to comment. We want to know if the rule is a) correct and b) useful.

The PPDM Association is the industry’s place for capturing knowledge about data management. Therefore, the Rules Library has a place for members to add advice about exceptions, the impact of an error and how to fix the data problem. And of course, we welcome your contributions of new rules!

*Certified professionals  
steward data ...*



*... to keep it useful  
for all stakeholders.*



[www.ppdm.org](http://www.ppdm.org)

**LUNCHEONS**

**OCTOBER 4, 2018**  
**MIDLAND DATA MANAGEMENT LUNCHEON & TRAINING**  
 Midland, TX, USA

**OCTOBER 10, 2018**  
**OKLAHOMA CITY DATA MANAGEMENT LUNCHEON**  
 Oklahoma City, OK, USA

**NOVEMBER 13, 2018**  
**DALLAS/FORT WORTH DATA MANAGEMENT LUNCHEON**  
 Dallas/Fort Worth, TX, USA

**NOVEMBER 14, 2018**  
**HOUSTON DATA MANAGEMENT LUNCHEON**  
 Houston, TX, USA

**DECEMBER 4, 2018**  
**TULSA DATA MANAGEMENT LUNCHEON**  
 Tulsa, OK, USA

**JANUARY 15, 2019**  
**HOUSTON DATA MANAGEMENT LUNCHEON**  
 Houston, TX, USA

**JANUARY 29, 2019**  
**DENVER DATA MANAGEMENT LUNCHEON**  
 Denver, CO, USA

**FEBRUARY 14, 2019**  
**OKLAHOMA CITY DATA MANAGEMENT LUNCHEON**  
 Oklahoma City, OK, USA

**WORKSHOPS, SYMPOSIA & EXPOS**

**OCTOBER 22 – 24, 2018**  
**CALGARY DATA MANAGEMENT SYMPOSIUM, TRADESHOW & AGM**  
 Calgary, AB, Canada

**OCTOBER 25 – 26, 2018**  
**MACHINE LEARNING WORKSHOP**  
 Calgary, AB, Canada

**NOVEMBER 7, 2018**  
**DENVER PETROLEUM DATA SYMPOSIUM**  
 Denver, CO, USA

**MARCH 5, 2019**  
**DALLAS/FORT WORTH DATA MANAGEMENT WORKSHOP**  
 Dallas/Fort Worth, TX, USA

**APRIL 9 – 10, 2019**  
**HOUSTON PROFESSIONAL PETROLEUM DATA EXPO**  
 Houston, TX, USA

**MAY 14, 2019**  
**OKLAHOMA CITY DATA MANAGEMENT WORKSHOP**  
 Oklahoma City, OK, USA

**CERTIFICATION - CERTIFIED PETROLEUM DATA ANALYST**

**NOVEMBER 14, 2018**  
**CPDA EXAM**  
 (Application Deadline October 3, 2018)

**FEBRUARY 27, 2019**  
**CPDA EXAM**  
 (Application Deadline January 16, 2019)

**MAY 22, 2019**  
**CPDA EXAM**  
 (Application Deadline April 10, 2019)

**AUGUST 28, 2019**  
**CPDA EXAM**  
 (Application Deadline July 17, 2019)

**NOVEMBER 13, 2019**  
**CPDA EXAM**  
 (Application Deadline October 2, 2019)

**ONLINE & PRIVATE TRAINING OPPORTUNITIES**

Online training courses are available year-round and are ideal for individuals looking to learn at their own pace. For an in-class experience, private training is now booking for 2019. Public training classes are planned for 2018 and 2019.

**OCTOBER 2018**

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

**NOVEMBER 2018**

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

**DECEMBER 2018**

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

**JANUARY 2019**

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

**FEBRUARY 2019**

S	M	T	W	T	F	S
						1
						2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28		

All dates subject to change.  
**VISIT PPDM.ORG FOR MORE INFORMATION**



Find us on Facebook  
 Follow Us @PPDMAssociation on Twitter  
 Join our PPDM Group on LinkedIn

**WE'VE GOT SOME OF THE BEST MINDS IN THE  
BUSINESS WORKING ON YOUR SUCCESS**

**EVOLVE, ALWAYS**  
That's geoLOGIC



Premium data, innovative software, integrated analytics for oil & gas.  
geoSCOUT | gDC [geologic.com](http://geologic.com)