PPDM™

Foundations

Journal of the Professional Petroleum Data Management Association

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

Volume 6 | Issue 2

1st Place Foundations Photo Contest Winner: *"Moraine Lake, Banff, Alberta" by Nicole Proseilo.*

The Value of Data Warehousing

The Shift (Page 4)

PLUS PHOTO CONTEST:

This issue's winners and how to enter (Page 20)

Well Lifecycle Data Management Solutions that Power the Enterprise

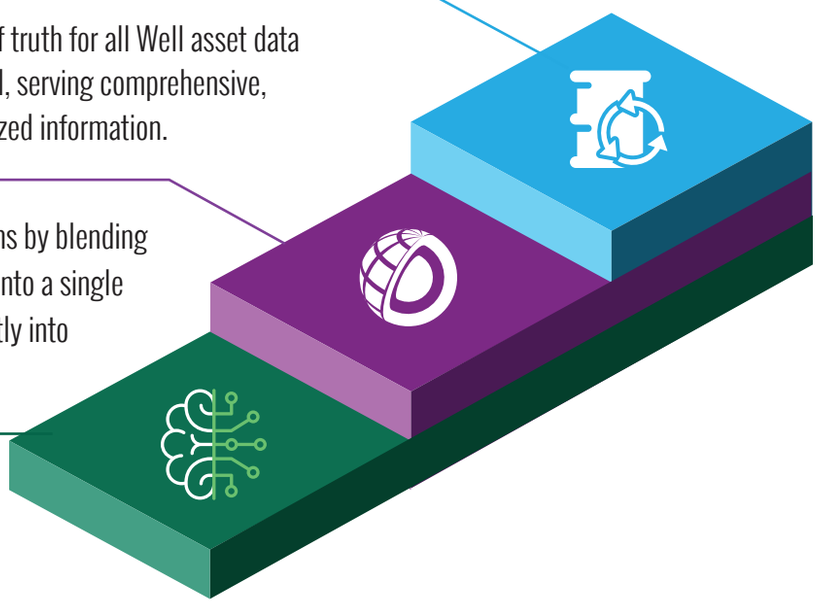
EnergyIQ
Information Certainty

We manage the data so you can use it.

Lifecycle Automation
Level 3
Maximize efficiency and analytics across the enterprise by automating rule-based workflows across critical applications.

Well Master
Level 2
Create a single source of truth for all Well asset data from concept to disposal, serving comprehensive, complete, and standardized information.

Competitive Intelligence
Level 1
Empower faster decisions by blending vendor and public data into a single source, streaming directly into analysis and workflows.



www.EnergyIQ.info | 303.790.0919 | info@energyiq.info

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
Margaret Barron, Matt Becker, Wade Brawley, Jim Crompton, Amii Bean, Melinda East, Dave Fisher, Guy Holmes, Peggy Loyd, Stewart Nelson, John Pomeroy, Francisco Sanchez

Editorial Assistance
Beci Carrington, Jim Crompton, Dave Fisher, Uwa Airhiavbere, Amii Bean

Graphics & Illustrations
Jasleen Virdi
Freepik

Graphic Design
DIGITAL PAGES INC.

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 6 | Issue 2

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 2019

Cover Feature

The Value of Data Warehousing

The Shift

By Amii Bean, EnerVest Operating

TRADITIONAL ENTERPRISE DATA WAREHOUSING

Before data warehousing, reporting and provisioning data from disparate systems was an expensive, time-consuming, and often futile attempt to satisfy an organization's need for actionable information (Raden, 2019). At an enterprise level, data warehousing provides an environment that mimics an umbrella over operational systems. It is designed in a way that supports data-based decisions, analytical reporting, data mining for statistical-based decisions, and even ad hoc querying at the end user level. The traditional data warehouse integrates data from legacy and transactional systems and other applications or sources. It also helps in resolving an ongoing issue of constantly pulling data out of operational systems by housing it in a space efficiently when extracted and transformed. Successful implementation of a data warehouse will allow an enterprise to convert stored data into good decision-making information.

WHAT DOES A DATA WAREHOUSE SOLVE?

The data warehouse was meant to solve a few underlying problems with disparate legacy operational systems. The first of three problems is **access to data throughout the enterprise**. "Being able to get access to the data at all is a breakthrough for many organizations" (Smith, 2017). Since the data is already in the operational systems, the data warehouse is not about creating or even moving the data. "The point of the data warehouse is to collect it in one place so that it is easy to answer questions quickly" (Smith, 2017). Which brings me to the second issue: the data warehouse is all about **driving data-based decisions** that answer business questions at the speed the business needs to stay competitive. Access and speed of data to the end user is great, but what about quality? The third issue is the ability to bring a **single version of truth** of an attribute to the end user. That data has to be correct and everyone has to agree on it and use that truth for data-based decisions. "Data from the various business units and departments is standardized [in a data warehouse] and the inconsistent nature of data from the unique source systems is removed" (BI-Insider, 2011). So, technically, the point of the data warehouse was not about the data. The data has always been there, collected and stored in siloed systems, but the data warehouse solved the problem of time consuming collection and transformation of that data into something useful.

BENEFITS OF DATA WAREHOUSING

Along with solving these three big issues, a data warehouse brings other benefits. With a data warehouse, "insights will be gained through improved information access" (BI-Insider, 2011), thereby enhancing business intelligence because "managers and executives will be freed from making their decisions based on limited data and their own gut feelings. Decisions that affect the strategy and operations of organizations will be based upon credible facts and will be backed up with evidence and actual organizational data. Moreover, decision makers will be better informed as they will be able to query actual data and will retrieve information based upon their personal needs. In addition, data warehouses and related business intelligence can be used and applied directly to business processes including marketing segmentation, inventory management, financial management, and sales" (BI-Insider, 2011).

Data warehouses also bring about some historical intelligence. "Data warehouses generally contain many years' worth of data that can neither be stored within nor reported from a transactional system.

Typically, transactional systems satisfy most operating reporting requirements for a given time-period but without the inclusion of historical data. In contrast, the data warehouse stores large amounts of historical data and can enable advanced business intelligence including time-period analysis, trend analysis, and trend prediction. The advantage of the data warehouse is that it allows for advanced reporting and analysis of multiple time-periods" (BI-Insider, 2011). These are strong enough reasons for management to jump on the data warehousing bandwagon, so what's the drawback?

CHALLENGES IN DATA WAREHOUSING

Is the Data Warehouse Dead?

Some database administrators (DBAs) noticed issues with data warehouses in the beginning. End users couldn't agree upon WHAT version of truth to use for a particular attribute. If new data being introduced into the data warehouse was not incorporated properly, it would sometimes "break" the warehouse, which took DBAs some time to fix, and that ends up being all the time. DBAs were stuck in a never-ending design phase as it seemed data warehouses were never finished (which could actually be less of a challenge, in that we can think of a data warehouse as ever-evolving, rather than not finished). That being said, some say the data warehouse is dead. "Like a dodo. Like Monty Python's parrot" (Howard, 2011). Some say it has no future in modern data management. There may be a few kinks to work through in a data warehouse, but it "is not dead… it is struggling. It is alive, but perhaps not entirely well. Big data, NoSQL, data science, self-service analytics, and demand for speed and agility all challenge legacy data warehousing. Traditional data warehousing...simply can't keep up with the demands of rapidly growing data volumes, processing workloads, and data analysis use cases. Data warehousing must evolve and adapt to fit with the realities of modern data management and to overcome the challenges of scalability and elasticity, data variety, data latency, adaptability, data silos and data science compatibility" (Wells, 2018).

Alive, But Not Well

The data warehouse is definitely alive, but not necessarily alive AND well. Several factors indicate the data warehouse is struggling. "Legacy data warehousing infrastructure addresses growth as a scale-up problem… [however] scaling up is inadequate for today's data volumes, processing workloads, and query rates" (Wells, 2018). There is also an issue with the variety in today's data. "Most data warehouses are implemented using relational database management systems…as relational technology was the predominant database technology of the day and most warehouse data was sourced from relational databases used by enterprise operational systems. The big data phenomenon radically expanded the variety of available data sources and the ways in which data is organized and structured" (Wells, 2018) to include unstructured, semi-structured, and multi-structured data. Data warehousing also has to be processed in an ETL (Extract, Transform, Load) manner with periods of time where the system needs to be refreshed specifically weekly, and even daily. "Batch processing is inherently latent. As the speed of business accelerates, data drives process automation; and digital transformation intensifies dependency on data. The modern data warehouse must be able to ingest and process data at the right frequency…" (Wells, 2018) for optimal efficiency.

"The early vision of data warehousing—a single place to go for integrated and trusted data—has clearly not become reality when 90% of companies operate two or more data warehouses. The reasons for multiple data warehouses are many, including mergers and acquisitions, independently developed departmental and line-of-business warehouses, geographically specialized warehouses for multi-national companies, and more. Regardless of the causes, multiple warehouses create the very data silos that

warehousing is intended to eliminate" (Wells, 2018). The challenge here is to design a data warehouse ecosystem that will adapt with the technological advances, encompass all varieties of data, be elastic and capable of scaling out, and coexist with data lakes so that the environment is compatible and can complement the data lake in order to prevent confusion, which is caused by conflicting information.

THE SHIFT IN DATA WAREHOUSING

What Does the Future Hold?

The times that a data warehouse is the single version of truth no longer works for modern data management. And the data warehouse as a hub for data integration and business intelligence applications doesn't work here either. "The BI and analytics world sees the data warehouse as just one of the many available data resources. So the question of fit becomes more complex" (Wells, 2017). In an article on the future of the data warehouse, Dave Wells, a research analyst for the Eckerson Group, stated that "the purpose of a data warehouse in modern data management architecture is to provide a repository of enterprise history that is integrated, subject-oriented, non-volatile, and time-variant. The warehouse characteristics that Inmon described so many years ago are still important. And historical data still matters. It doesn't meet all analytic needs, and it doesn't have the 'shiny object' appeal of real-time and streaming data. But it is the essence of time-series analysis that is at the core of decision support and performance management." An enterprise's data management ecosystem has a place for data warehouses as components within an enterprise data hub.

"They exist together with MDM [Master Data Management], ODS [Operational Data Systems], and portions of the data lake as a collection of data that is curated, profiled, and trusted for enterprise reporting and analysis" (Wells, 2017).

Here for the Long Haul

So much time has been invested in the design and function of the data warehouse that it is here to stay. Business units and functions all depend on the data warehouse now, "but sustainability demands that we rethink the data warehouse. Data warehouse architecture can no longer stand alone. We must think purpose, placement, and positioning of the data warehouse in broader data management architecture. Architecture, of course, is only the beginning. The data warehouse is alive but it faces many challenges. It doesn't scale well, it has performance bottlenecks, it can be difficult to change, and it doesn't work well for big data. In a future of data warehouse modernization, we'll need to consider cloud data warehousing, data warehousing with Hadoop, data warehouse automation, as well as architectural modernization" (Wells, 2017).

MODERNIZING THE WAREHOUSE

The Drive for Something New

We know data warehouses have been around for years, and we've walked through the values they provide and the challenges they face in modern data management. So what drives the need to modernize the data warehouse? The challenges in legacy enterprise data warehouses are the drivers: the need for flexibility; the need for streamlined architecture, for example, data warehouse as a service, or data warehousing in the cloud; the need for agnostic data; and the ability to deploy "a new kind of data warehousing that needs to support newer BI deployments to keep up with customer demand. The main factors that drive development and deployment of new data warehouses are being agile, leveraging the cloud and the next generation of data as it relates to real-time data, streaming data and data from IoT devices" (Thomas, 2018). So how is this accomplished? "When it comes to big data, leveraging the latest technology or processes is the key to efficiency. [D]ata transformation is moved to the end in the traditional ETL process, which is considered the modern standard now. So instead of having the data extracted, transformed, and loaded directly into the data warehouse, in the case of ELT (Extract, Load, Transform), once the data is extracted it is directly ingested. Later, it is then transformed on read. The big advantage of this process is its flexibility via a quick code change on the view versus making significant changes to your transformation process" (Thomas, 2018). A company must be able to adapt to the ever-changing technology. "Specifically, focus on being agile, have a cloud adoption strategy, and partner with an industry ETL expert that knows innovative processes as well as know your business objectives" (Thomas, 2018).

The Cloud Revolution

Being flexible with an ETL or ELT process isn't the only driver for modernizing the data warehouse. Most of today's big data revolution is the advancement of cloud data storage and cloud computing. "The cloud transformed the notion of what's possible when architecting and building the data warehouse" (Snowflake, 2018). And this is the reason that the data warehouse needs to evolve to keep up with technology demand. "A crucial part of that evolution is the need to develop a new architecture that can support unlimited growth in data volume, workload intensity, and concurrency. Another critical component is support for diverse data: traditional, relational data, and semi-structured data. Ease of use matters. Without it, all potential data users can't have access to effective data analytics. 'Why not cloud?' has become the default question" (Snowflake, 2018), and rightfully so. "History reveals [that] the modern data warehouse must leverage the architecture and affordability of the cloud, the flexibility of NoSQL technology [along with SQL], and the power of the traditional warehouse" (Snowflake, 2018). F

REFERENCES

Amazon Web Services. (2018). What are NoSQL databases? Accessed July 26, 2018, from https://aws.amazon.com/nosql/

Barton, N. (2018, April 13). Cloud: The future of the data warehouse. Accessed May 20, 2019, from https://www.dataversity.net/cloud-future-data-warehouse/#

BI-Insider.Com. (2011, July 31). Benefits of data warehousing. Accessed May 30, 2019, from https://bi-insider.com/portfolio/benefits-of-a-data-warehouse/

Dailey, D. (2019, March 11). The enterprise data warehouse of the future. Accessed May 20, 2019 from https://www.ibmbigdatahub.com/blog/enterprise-data-warehouse-future

Howard, P. (2011, June 29), The EDW is Dead. Accessed May 20, 2019, from https://www.bloorresearch.com/2011/06/edw-dead/

Raden, N. (2019, January 15). The data warehouse is dead – Long live the data warehouse. Accessed May 20, 2019, from https://diginomica.com/the-data-warehouse-is-dead-long-live-the-data-warehouse

Ryan, J. (2018, July 20). The future of data warehousing. Accessed May 30, 2019, from https://dzone.com/articles/the-ideal-warehouse-architecture-its-in-the-cloud

Singh, A., & Coyne, B. (2018, September 12). The future of data warehousing. Accessed May 20, 2019, from https://www.oreilly.com/ideas/the-future-of-data-warehousing

Smith, S. J. (2017, July 19). The demise of the data warehouse. Accessed May 30, 2019, from https://www.eckerson.com/articles/the-demise-of-the-data-warehouse

Snowflake. (2018, April 15). The past, present, and future of data warehousing. Accessed May 30, 2019, from https://nl.devoteam.com/wp-content/uploads/sites/15/2018/04/historyofdw-final-1.pdf

Thomas, B. (2018, January 2). The data warehouse in 2018. Accessed May 20, 2019 from https://www.cio.com/article/3245386/the-data-warehouse-in-2018.html

Van Loon, R. (2019). What is the future of data warehousing? Accessed May 20, 2019. https://mapr.com/blog/what-future-data-warehousing/

Wells, D. (2017, October 10). The future of the data warehouse. Accessed May 20, 2019, from https://www.eckerson.com/articles/the-future-of-the-data-warehouse

Wells, D. (2018, September 30). The future of data warehousing. Accessed May 20, 2019, from https://www.eckerson.com/register?content=the-future-of-data-warehousing-integrating-with-data-lakes-cloud-and-self-service

Wells, D. (2019, August 14). The data warehouse is still alive. Accessed May 30, 2019, from https://www.eckerson.com/articles/counterpoint-the-data-warehouse-is-still-alive

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.

Feature

Leading Your Organization to Better Data Governance

By Wade Brawley, Land Information Services

The terms ***data,*** *information*, and *knowledge* are not synonymous. Data is simple, one-dimensional, isolated, and unintuitive – merely building blocks for information. Isolated bits of data are not very useful until they are woven together into meaningful information that helps the customer to make a decision. So, it is extremely important for data to have accuracy and integrity in order to make a successful transformation to information. Any incorrect data component may lead to inaccurate information, which can lead to an incorrect assessment and, possibly, a devastatingly unfortunate decision. If you're still with me, then you understand that the next iteration is to build knowledge from information. It's knowledge that enables companies to surpass their competition by building predictive models to aid better investment strategies and optimum field operations.

If we have successfully established that data is the most granular component of the supply chain that contributes to a company's knowledge and success, then I believe we have underscored the importance of data responsibility. And who is responsible? Following is a good summation of everyone's role in data, no matter what department:

- **Data producers** create, update, and occasionally delete data, typically as a part of carrying out their primary business function.
- **Data stewards** have overall responsibility for ensuring data is managed in a way that serves the full value chain and supports all data roles across the division. A data steward audits and trains proper utilization of systems that capture data input by data producers.
- **Data administrators** carry out day-to-day coordination, monitoring, and remediation activities, and assist in developing division and project-level data policy in a manner that serves the full value chain. Data administrators are custodians who prepare data for use by others.
- **Data consumers** would define most of us, particularly management level personnel, who consume data in order to carry out operational or decision-making activities.
- **Subject matter experts** are professionals who have been in the trenches at various levels and are consulted with specialized knowledge concerning key data types and processes.

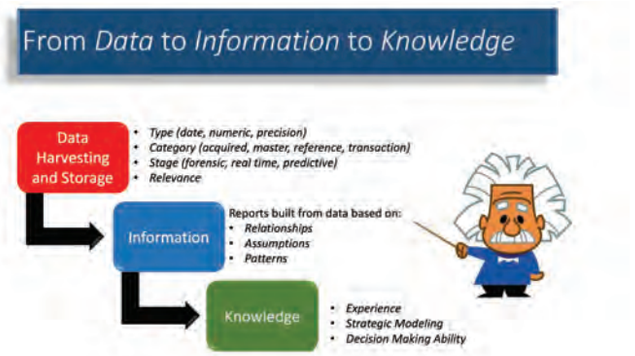


Figure 1: From Data To Information to Knowledge

Did you identify with at least one of these roles?

This brings us to the distinction between *data management* and *data governance*. **Data management** is the practice of organizing your data resources so that they are accessible, reliable, and timely whenever users call on them. Data management may include:

- Data stewardship.
- Data architecture.
- Data quality management.
- Data warehousing.
- Business intelligence and analytics.
- Metadata management.

If data management is the logistics of data, **data governance** is the strategy, security, and policy of data. Data governance is the setting and managing of policies and standards to ensure quality and integrity as data moves through

its enterprise value chain. It defines expectations, sets authority, and monitors performance. Data governance should feel bigger and more encompassing than data management and, for larger organizations, is the logical next step. In smaller organizations, data management and data governance are usually blended into one group, or even one person. Data governance should permeate the organization and encompass the following initiatives:

- Acknowledging the value of your company's data.
- Standardizing data systems, policies, and procedures.
- Promoting transparency.
- Ensuring correct regulation and compliance procedures.
- Helping to solve issues with data.
- Establishing training and education around data.
- Increasing enterprise revenue overall through accuracy and integrity, leading to better decisions.

DATA MANAGEMENT AND DATA GOVERNANCE ILLUSTRATED

Personnel turnover creates vulnerability in data quality. One aspect is the risk of disgruntled employees attempting to take valuable information to their next job. Data governance should have a policy and standards that protect data from being vulnerable to theft. In addition, personnel turnover impacts a company's data through misinformation among data producers. As a person leaves for another role in the same company or outside the company, they usually train their replacement. It is possible for bad habits to be perpetuated, or for the new data producer to invent their own rules on how and when data fields are populated. Data governance can prevent this event by building "guard rails" for practices and procedures so data quality is not vulnerable to natural shifts in personnel and management. Some of the guard rails can be programmed into the data producer's database, while other guard rails must exist in written documentation supporting the proper use of the data producer's database.

If you are faced with justifying the implementation of a data governance program for your company, the following tips may be helpful:

- Acknowledge that there are stages of data (forensic, real-time, and predictive) and that these stages are relevant to various types of data: master data, reference data, and transactional data.
- Illustrate data sources; the process of identifying all data sources will quickly demonstrate the complexity of your data landscape and identify redundancies where data components can conflict. The result is a useful reference for data administrators, as well as a compelling argument to management for the need for data governance.
- Agree to yield to other team members. A democracy is easier to lead than a dictatorship. If you do not sufficiently engage stakeholders from the beginning, you risk failure to buy-in and long-term commitment.
- Do not try to change history. Ignoring the company narrative simply because you want the initiative to carry your exclusive brand is self-serving. While change may be paramount, it's important to evaluate the benefits of current people, processes, and technology.
- Establish a democratic process with checks and balances so that one person or group cannot railroad their solution. The solution must make sense to everyone. It's imperative to include data stewards and secure endorsements and support.
- Identify champions among data producers, stewards, and consumers. These champions will help support your initiative for a successful implementation and an evergreen process.
- Present real examples of costly data errors:
  - Over- or underestimating reserves.
  - Double counting production.
  - Misallocating production.
  - Untimely revenue received.
  - Penalties paid for late disbursement.
  - Losing a lease.
  - Paying too much for an asset.
  - Under-evaluating a proposed asset and losing the deal.
  - Drilling a dry hole.

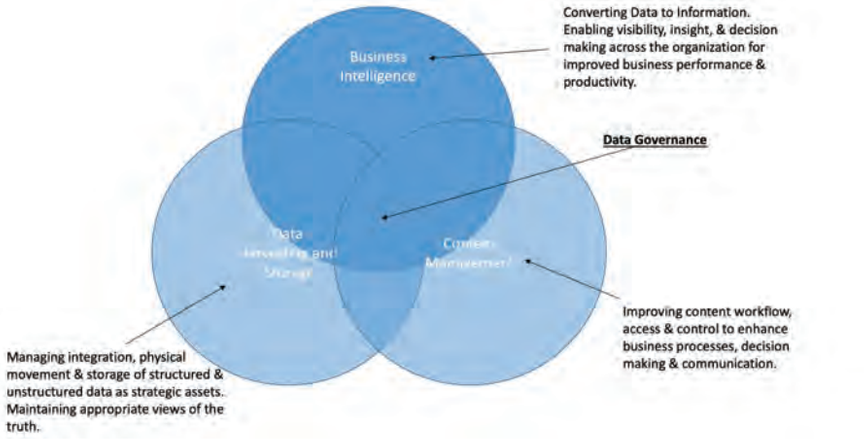


Figure 2: Data Management and Data Governance Illustrated

At this point you should be able to obtain agreement from management that more accurate and timely data will reduce risks and costs, facilitate more accurate and faster decisions, and lay the foundation for establishing key performance indicators (KPIs) for greater efficiency and accuracy.

About the Author

Wade Brawley is CEO for Land Information Services, which specializes in workflow solutions through its software platform – *LandVantage*.

Innumeracy and Math Abuse

or, How to Lie with Data

By Jim Crompton, Reflections Data Consulting

With great power comes great responsibility. The amount of Big Data that is now accessible and the powerful analytical tools and algorithms that are now widely available reinforces that data brings great power. Advanced analytics also carries the responsibility to inform and not mislead. How about adopting a new creed in analytics of "first, do no harm" (my apologies to Hippocrates)?

DATA VISUALIZATION

Data visualization is one of my favorite topics these days. Visualization is an art form. An experienced practitioner can perform magic while most of us are just struggling to get the plot to look about right. A clever image that condenses a complex problem with lots of data into a clear story with an obvious recommendation is a work of art. Just as telling a story with data (and graphs and images) is the critical output of a good analysis, these same tools can turn to the dark side and mislead, misinterpret, and fool the unsuspecting audience. Beyond the problem of intentional deception is unintentional bias. In this article, I am not going to dwell on the bad guys as much as on the unintentional mistakes that all of us can make with advanced analytics.

Misleading graphs and chart abuse: it can be amazing how the choice of axis on a two- or three-dimensional plot can mislead or confuse the viewer. The shape of a histogram changes when the size of the bins is changed. A bin width with too narrow an interval range (for example, one that includes just 10 variables as opposed to 500) may result in too much variation, causing the viewer to "miss the forest for the trees."

There are as many great visualizations as there are ones that confuse (too many variables plotted) and mislead (poor choice of axis ranges). Color can be a critical way to emphasize key insights, but also a way to confuse (e.g., red/green color blindness).

BIAS

Bias is any systematic failure of a sample to represent its population. Sampling methods, by their nature, tend to over- or underemphasize some characteristics. The best way to avoid bias is to pick samples at random. Selecting a truly random sample data set before starting your analysis is one of the biggest challenges in statistics. But how often do we really think about this issue? How often do we just grab all the data available and start turning the analytical crank to build a prediction model?

The assumption of a normal distribution is critical in many traditional statistical techniques. Figure 1 shows how bias may affect the calculated mean that is supposed to represent the entire population.

Selection bias involves individuals being more likely to be selected for study than others. If I survey my friends and 75% of them like the Denver Broncos and if I extend this observation to the whole population of American football spectators, I have made the error of selection bias.

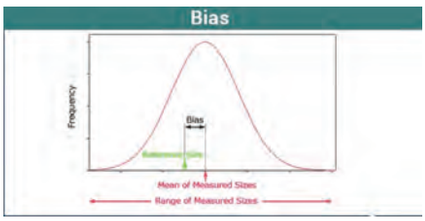


Figure 1: Bias

Bias through placement: We tend to focus on news on the front page of the paper or on websites and ignore the stories on the last page (unless we are looking for our name in the obituaries).

Confirmation bias: Cognitive dissonance is the anxiety we get when the real world does not fit with our world view. We may search out things that agree with us while ignoring conflicting data. We have become a culture of first stating our opinion and then looking for data to support that opinion. The old saying that "you are allowed your own opinion but not your own data" seems to be failing the test of popular application.

Anchoring: Our values are often set by imprints in our minds, which we then use as mental reference points when making decisions. For example, I grew up in the Rocky Mountains and have lived here most of my life. But when I was transferred to Houston, and first saw the Gulf of Mexico, my bias was to complain about the heat and humidity as well as the hurricanes. My oldest daughter grew up in Houston and now lives in Florida. When she comes to visit us, her complaints are about the

snow and cold weather, as well as lack of a good beach. That is anchoring.

Spectrum bias arises from evaluating diagnostic tests on biased samples, leading to an overestimate of the sensitivity and specificity of the test.

Omitted-variable bias appears in estimates of parameters in a regression analysis when the assumed specification omits an independent variable that should be in the model. Data prep steps (filtering) can have a lot to do with what data is accepted from a noisy data sample.

MORE PITFALLS

Sampling theory is concerned with the selection of a subset of individuals (e.g., my Denver football friends) to estimate characteristics of the whole population. Each observation measures one or more properties of observable bodies distinguished as independent objects or individuals. Done correctly, sampling reduces the size of the data set for analysis. Done incorrectly, your data set does not represent the original population; it is biased.

Misunderstanding of probabilities and **ratio anxiety** can also affect your analysis. For example, if original test scores average 100, but then fall by 60%, but rise by 70% when the test is retaken, are you better off? Think about this one carefully... And the list goes on and on.

Correlation does not imply causation. Many statistical methods are powerful tools for calculating correlation between variables. The correlation coefficient expresses the "goodness of fit," but taken too far, correlations can be mistaken for root causes. Check out this website for some very funny examples: http://www.tylervigen.com/spurious-correlations.

Executives often overlook potential perils ("We're not using AI in anything that could blow up, like self-driving cars" or producing wells) or overestimate an organization's risk mitigation capabilities ("We've been doing analytics for a long time, so we already have the right controls in place, and our practices are in line with those of our industry peers"). It's also common for leaders to lump AI risks with others owned by specialists in the IT and analytics organizations ("I trust my technical team; they're doing everything possible to protect our customers, operations, and our company").

Data quality really matters. As the amount of data (structured, unstructured, transactional, social media, field sensors and mobile devices, videos, documents, and drawings) from diverse sources has increased, steps such as ingesting, sorting, and integrating have become increasingly difficult. It's easy to fall prey to pitfalls such as inadvertently using or revealing sensitive information hidden among anonymized data (I have already covered the sampling bias pitfalls).

Artificial intelligence methods generate both benefits and business value. AI is also giving rise to a host of unwanted, and sometimes serious, consequences. The most visible ones — privacy violations, discrimination, accidents, and manipulation of political systems — are more than enough to prompt caution. More concerning still are the consequences not yet known or experienced. Disastrous repercussions—the loss of human life if an AI medical algorithm goes wrong, or the compromise of national security if an adversary feeds disinformation to a military AI system — are possible and so are significant challenges for organizations, from reputational damage and revenue losses to regulatory backlash, criminal investigation, and diminished public trust. As one example, consider how the big internet companies are wrestling with the issue of trust with their users' personal data and how that could impact future demand for their platform.

Another example, one that has been in the news recently, is facial recognition software. It detects faces in an image, separates each face from the background, normalizes the position of the face, and the result is thrown into a neural net for feature discovery and classification. The software then compares the face to faces in a database to see if there's a match. When the algorithms (such as the open source favorite OpenFace) are trained from a biased data set with supervised learning methods, the prediction model will do well with some images (those similar to the test data set) and poorly with others with uncertain outcomes.

Getting a prediction model right depends on a random data set to train it. That is not as easy as it sounds. There are always problems building a training set. The oil and gas industry, especially drilling, is calling for more sharing of data to address this point. How an analyst extracts features and attributes does matter. Homogenous development teams can produce homogenous results (and most development teams consist of young males who are predominantly white or Asian). No wonder many algorithms have trouble with images from other races and genders. Even extensive tests of an algorithm on non-diverse (not random) data sets will not result in accuracy from a wider set of images. Maybe the problem is using lab-quality test data from sensors and then having an algorithm fail on the noisy actual field sensor inputs.

AI models can create problems when they deliver biased results, become unstable, or yield conclusions for which there is no actionable recourse for those affected by its decisions (such as someone denied a loan with no knowledge of what they could do to reverse the decision or a stuck pipe condition that is not recognized by an algorithm not trained on these conditions).

BE CAREFUL OUT THERE!

Whether you are the one building the model or the decision-maker being shown the analysis, it is well worth taking the extra time to build a training set with a representative distribution of all possible varieties and randomness. This is yet another case where the data is important, not just the technology. F

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.

Buzzwords and Technology Basics in 2019

Internet of Things (IoT)

By Guy Holmes, Tape Ark

Last issue, my column covered Web 4.0. I started with Web 4.0 because all the other buzzwords that we will cover form a part of the Web 4.0 ecosystem. Web 4.0 is, in many ways, both a tool in itself and an enabler of many other tools – a symbiotic relationship of technology to technology, and technology to human integration built to enable a better future.

This issue, I want to extend that discussion of technology to technology integration by covering Internet of Things (IoT). As a bit of background, IoT was minted in 1999 by Kevin Ashton, who worked in the supply chain area of Proctor and Gamble. Kevin wanted to get the company involved in the advancements of RFID (radio frequency identification) and needed a way to express the concepts to his management team.

I too played around with RFID many years back and was so excited by its potential. RFID stood to allow significant improvements in inventory management, essentially by replacing barcodes on items. RFID "tags" (stickers with a small chip inside them) would allow the inventory of items by scanning an entire pallet's contents with a single scan. The RFID tags would then transmit their presence back to the RFID reader and the reader would be able to collect all the signals and essentially perform a count and inventory of the pallet in a matter of seconds.

There is an old story of a frozen food company that would send out pallets of frozen peas to grocery stores. Each pallet was supposed to contain 400 bags of peas. However, knowing that no one would ever count the number of bags of peas in each supermarket when the inventory arrived, they started to leave 10 bags of peas off each pallet to increase their own profits. The RFID tag stood to change the evil ways of the frozen pea industry forever and eventually it did.

Since the time of RFID introduction, IoT has come a long way and moved so far away from RFID that most would not recognize it as the same concept.

To best understand IoT, we do not need to look up the definition of what an IoT sensor is or the technology behind it. It is better to look at use cases and the things we use every day.

So, where is IoT used today?

1. The Smart House – anyone doing renovations or building a new house knows that a tonne of their options available directly involve the use of IoT. Be it smart thermostats, security cameras, or power plugs that can be activated, viewed, and controlled from your phone, these devices are part of the growing IoT ecosystem.
2. Smart Agriculture – farmers are increasing their reliance on sensors that can check soil composition, moisture, or signs of disease. When combined with drone technology to fly over crops using imagery and collecting additional data, the agricultural world is seeing major improvements in crop yield and profitability.
3. Wearables – wearables are a class of IoT technology on their own and the market is simply gobbling this technology up at each new stage of its growth. Everything from Fitbits to cardiac rhythm and UV sunburn monitors fall into the wearable IoT category. In fact, this is one of the fastest growing IoT markets in the world.

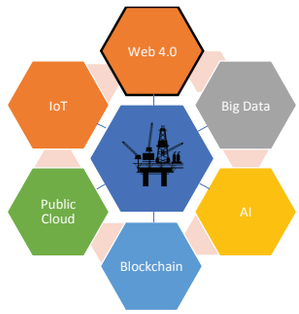


Figure 1: Trending Technology Buzzwords.

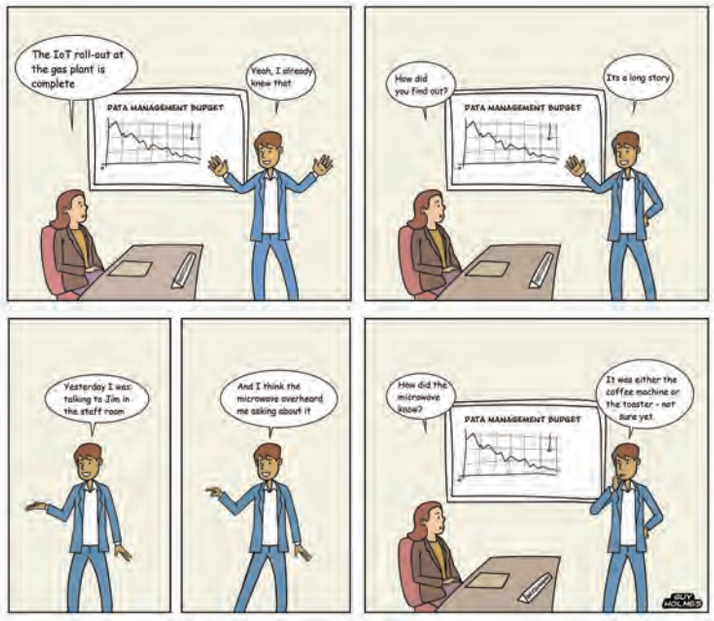
The growth of IoT has been massive, not only in consumer areas such as health and smart houses, but also as a part of the Web 4.0 industrial revolution and its use in heavy industry, such as oil and gas. In fact, according to the McKinsey Global Institute, IoT has the potential to make an $11 trillion impact by 2025, and a significant amount of that in the oil and gas sector. One of the things about IoT is that the best implementations of IoT are often not highly visible, but are definitely highly impactful.

So, how does IoT influence our daily activities in the oil and gas sector if we don't see it? The better question might be, how does it not influence our sector? A great example of IoT in the oil sector is an Australian major who recently developed their own sensors, which they plan to deploy to gas plants in Australia. The sensors will

collect a wealth of information, including temperatures, flow rates, and vibration activity and transmit this information in real time back to its monitoring bases.

In case you missed it, if you are a member of PPDM and obsessed with data, then get ready; this data will be flowing in real time. That's right – data will flow not days or weeks after an event, but will be extremely dynamic, continually growing, and relentless in its flow. In fact, the effects of this trend on PPDM will likely result in the need for rapid change in its training of data managers, changes to its data model, and the need for completely new types of data to be captured.

The data being captured now in the oil and gas sector is moving from present and past data content to predictive data types for events that have not even occurred, particularly in predictive maintenance areas. The aim is to know what is happening right this second as well as what is likely to happen in a few weeks, months, or years. In the next issue, we can use the IoT example to better understand the use cases for Big Data and AI as both of these technologies will be consuming the Web 4.0 IoT driven data deluge that is heading our way. F



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 another to founding a leading global data management company.

Feature

Wrestling the Data Quality Bull

Take Data by the Horns and Tame It

By Matt Becker, Sullexis

Every July, the city of Pamplona hosts the Running of the Bulls event as part of their summertime festivals. The origin of the festival stems from several centuries ago when the young cowboys, who were responsible for the care of the bulls, used the bull run event to show off their bravado and test their skills in herding the bulls to the central market. While many of us would look at this with a mix of awe, fear, and trepidation, there is something to be said about the proverbial grabbing the bull by the horns and taming it. How many times have we been told to face our fears, tackle the problem head-on, and just go for it? Well, like those cowboys of old herding their cattle, wrangling the bull that is data quality can be a daunting task, but this particular bull we no longer need to fear.

Figure 1: Running of the Bulls

What is data quality…what does it mean? Data quality means being able to provide reliable, trustworthy, and fit-for-purpose data to help enable data-driven decisions and optimize your operational processes. Data quality enables you to answer questions at any given moment, such as: how many wells are actively being drilled in your operating regions, or what work orders are currently being processed, or are your maintenance and reliability crews on schedule, or can you accurately see your field-reported costs compared to your AFEs and forecasted costs?

DATA QUALITY FRAMEWORK

With the increased adoption of big data platforms and cloud-based storage, and a greater need to access, interrogate, and examine the data contained within them, many organizations recognize how critical data quality is to their day-to-day operations. So how do we start to fence in the data quality bull? We utilize a modern-day data quality framework to help maintain and support good data quality across the various upstream functions. By following a straightforward, data stewardship-based approach, you can create a data quality management framework that focuses on integrating processes, technologies, and resources to maximize quality and value from data while reducing cost and risk.

There are three key areas that an effective data quality framework needs to incorporate, as depicted in Figure 2:

1. **People (Who)**
   Identifying who your data users are, addressing communication and training needs, and ultimately empowering and enabling your users to own and manage their data quality.
2. **Process (What)**
   Defining the data quality processes, how those fit across the well lifecycle, and how to make data quality a sustainable and ongoing activity that is imbedded into core operational functions.
3. **Technology (How)**
   Implementing and leveraging a componentized modern-data architecture that can scale and fit to current data quality needs and future requirements as well as provide flexibility to technology and industry changes.

USER-DRIVEN DATA QUALITY STAGES

Using the data quality framework as a guide, it's important to put the control and ownership for managing data quality into the hands of the user. Because of their daily interaction with the data and the subject matter knowledge they possess, users need the ability to address data quality issues in an efficient manner. Leveraging the

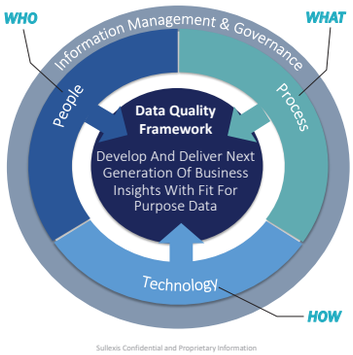


Figure 2: Data Quality Framework

data quality stages allows users to define and manage their own data quality rules, data corrections, and data monitoring in a step-by-step approach, building on and improving data quality as they progress through the stages. The added benefit is that users don't have to make repeated requests to IT for every new change or update needed. IT, in turn, can focus on infrastructure and technical capability enablement without having to devote resources to solution development. This is an opportunity for the business and IT teams to collaborate on data quality efforts by more efficiently utilizing their combined resources. As noted in Figure 3, there should be four key data quality stages: Discover, Define, Implement, and Monitor.

DATA QUALITY DIMENSIONS

As you move through the data quality stages, as shown in Figure 3, you can start to herd and guide your data quality bull to where you need it to go. Your users can focus on understanding their data quality by applying a set of data quality dimension standards to assessing data content, structure, and anomalies. These dimensions can be applied, at the user's discretion, based on their specific requirements. As you progress through the stages, identifying areas of data gaps and errors, you can monitor your data quality by building out a scorecard based on the data quality insight gained. By capturing and analyzing the data quality trends found in the data set(s), quality can be assessed across a set of

Figure 3: Data Quality Stages

data quality dimensions based on certain expectations about how the information describes a real thing (object or event):

1. Completeness:
   What data is missing or unusable?
2. Conformity:
   What data is stored in non-standard formats?
3. Consistency:
   What data values give conflicting information?
4. Duplicates:
   What data records or attributes are redundant?
5. Integrity:
   What data is not referenced or otherwise compromised?
6. Accuracy:
   What data is incorrect or out of date?

As the PPDM Data Rules working team has noted, in the information management world there is no standard or consensus on what quality dimensions are necessary or how the terms are defined. However, the six dimensions outlined provide a very good baseline for profiling, improving, and monitoring data quality within your organization.

HOW TO GET STARTED – KEEP IT SIMPLE WITH A DATA QUALITY(DQ) PROFILE

Too many times, organizations try to do too much all at once when starting data quality projects. They try to wrangle with too many data quality bulls at the same time. Data quality is a journey that will take time and commitment, but that doesn't mean results cannot be seen quickly.

The key is to start small. Start in the "01 Discover" stage by creating an initial data profile for a small data set (pick one or two data quality bulls) and keeping the scope tight for the initial data set effort by following these simple steps for a data quality (DQ) profile.

1. Establish the DQ profile.
   a. Identify a small sample data set.
   b. Extract and format the data (in Excel, for example).
   c. Conduct a set of basic data profiling techniques quickly and thoroughly.
2. Examine the DQ profile.
   a. Analyze the results of the data profile.
   b. Identify data quality issues.
   c. Provide recommendations.
3. Act on the DQ profile.
   a. Act on results through data cleansing efforts.
   b. Adjust business rules via better understanding of the data.

After spending time in the Discover stage, you will be able to prioritize the data quality issues and organize your efforts to move through the other Data Quality Stages.

KEEP CALM AND TAME THE BULL

This Data Quality Framework and stages will provide a strategy to execute, a remediation approach to monitor and track, and a set of business and technical processes to align users. By leveraging these leading data management practices, tame that wild bull called data quality and provide your people and your key operational systems with a set of data that is accessible, reliable, and trustworthy.

About the Author

Matt Becker serves as the Managing Director of Sullexis' Enterprise Data Strategy and Solutions practice.

PPDM™

BUILDING DATA PROFESSIONS

Communities work together...

...to create International Petroleum Data Standards...

...that are used to support a trusted and useful discipline...

...that benefits industry.

COMMUNITY

MEMBERSHIP

14,000+ Connections

114 Corporate and Contributing Members

3,144 Individual Members

IN

WORK FOR

934 Companies

58 Countries

214 Volunteers

EVENTS

7,200+ Subscribers

Record Attendance

2,480 Total Attendees

PPDM WORLDWIDE EVENTS

37 PPDM Events

9 Leadership Teams

81 Volunteers

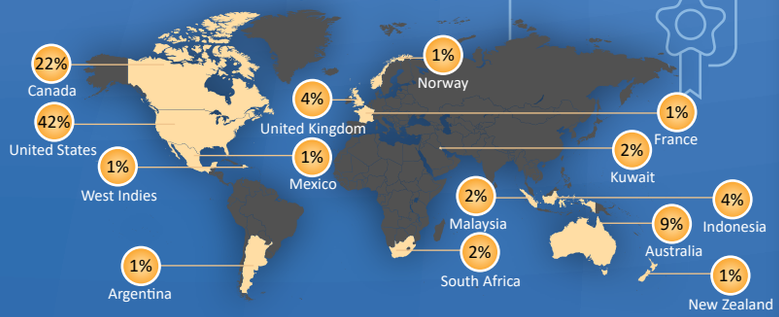
13 Cities

399 Houston 2019 Expo Attendees

PROFESSIONAL DEVELOPMENT

CERTIFICATION

CPDA Certified by Country



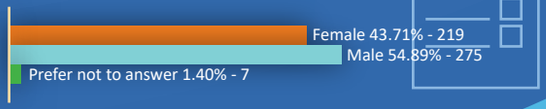
JOB FAMILIES

Level 1	Level 2	Level 3	Level 4
Corporate / IT	Upstream	Strategy / Governance	
		Standards / Policy	
		Projects	
		Digital Transformation	
	Midstream / Downstream		
Upstream	Subsurface	Geoscience	Wells
			Seismic
		Engineering	Well (D&C)
			Reservoir
			Production
		Geospatial	
	Surface	Facilities Engineering	Engineering
			Geospatial
	Land		
	Finance		
Midstream			
Downstream			
New Energy			

PD CATALOGUE LAUNCH

- **2** Data Management
- **2** Business Process and Life Cycle
- **2** Managing Location and Spatial Information
- **4** Industry Standards: Non-technical
- **PPDM Training**
- **3** Well Identification and Location Systems
- **17** Public Petroleum Data Model 3.9
- Postsecondary Offerings

COMPENSATION SURVEY RESPONSES: **501**



INTERNATIONAL PETROLEUM DATA STANDARDS (IPDS)

PPDM HELPS SOLVE KEY DATA CHALLENGES WITH ITS POWERFUL, COMMUNITY DRIVEN, INDUSTRY BASED METHODOLOGY

"The PPDM Way"

Competence Capability 1
PPDM Certification
CPDA
PPDM Training
30+ Classes

Data Trusted 2
PPDM Rules Library
3,000+ Rules
rules.PPDM.org

Clear Asset Identification 3
PPDM Well Identification
Global Best Practices
US Well ID
Canada Well ID

Structures Compatible 4
PPDM Data Model
60+ Subject Areas
PPDM 3.09

Semantic Clarity 5
PPDM Semantics
What is a Well
What is a Completion
Well Status

Clear Data Relationships 6
PPDM IPDS
Best Practices

Feature

Can Master Data Management Be Agile

And Can PPDM Help?

By Melinda East, Equinor & Stewart Nelson, Infosys Consulting

Is it possible to implement Master Data Management (MDM) technology and concepts in today's energy environment by utilizing work through PPDM and the Agile project framework? Learn of the trials and tribulations throughout this accelerated journey.

More than ever, operators are embracing digital solutions as one of the significant keys to differentiation and competitive advantage. Analytics, artificial intelligence, remote operations, robotics, and many other innovations hold great promise for safer and more profitable operations.

Demands for these solutions rapidly turn into deadlines and expectations. Vendors queue up with the newest and shiniest solution to revolutionize the performance of your well or asset, or to optimize your workflow; they even profess to predict behavior. The common denominator in many of these innovations is the fuel – the data and information flowing in this digitally integrated ecosystem. This article explores a solution approach and the role PPDM could play in enabling success.

What does it take to build foundational data aspects to drive the ambitions of the company and evaluate or leverage the multitude of innovational opportunities? As with all journeys, it takes the first step and, in the case of this operator, the first step was to recognize the value of a data management capability versus the traditional asset-by-asset approach to information management. It took a great deal of time, but this process developed enough trust to continue progressing. Over this time, the users became the front end of the curation of the data, which is critical for long-term adoption. The next step, multi-domain MDM, was significant and the approach had to align with the needs of the business and deliver results quickly. The objective from management was stated clearly. Deliver quick results with reliable data that is seen across the value chain while positively impacting the agendas of a digital future and doing so with great flexibility to accommodate for adjustment.

Many data professionals have faced the MDM challenge of laying out the master relationships between data and systems, and data and entities, and of automating those relationships. It can take a great deal of time; a long-cycle engagement would not be welcomed. The team chose an Agile approach to introduce multi-domain MDM in the cloud. While this approach was met with apprehension, it was allowed to progress.

The approach required segmenting the key master data elements and defining accelerators that would lend themselves to the rapid prototyping native to an Agile method. The PPDM model, definitions, and all of the artifacts created the perfect platform for successful rapid requirements (backlog) development. In this use case, an MDM application was thoughtfully selected prior to initiation of the full business engagement, so this acted as another accelerator.

Another key to the pace, as well as to the overall success of the Agile approach to MDM, was the assignment of executive and commitment of the product owners for each data type. These key contributions and accelerators were foundational to the success of this MDM program.

Based on the parameters set, a truly Agile approach was adopted and communicated to participants and stakeholders. Many

not directly aligned with the project team could see only the output and rapid results, but not see or fully experience the people challenge created. To counteract this, experienced data, operations, and technical professionals were assembled from the operator, MDM software provider, and implementation consultant. A level of experience with MDM or Agile existed, but there was very little, if any, MDM experience using an Agile model. Managing non-core team members or peripheral participants required education and communication to help develop an understanding of the frequency of involvement and what they should expect at the end of each sprint or major release. A unique outcome resulted from this high frequency engagement model. We noticed small but frequent expectations verification provided users with the opportunity to look at tangible and non-tangible problems over time and consider the options of design aspects of the solution. User "buy in" was more genuine because their input was considered and calculated (and recalculated a few times in many cases).

Rapid pace requires refined coordination, courageous communication, and a level of accountability across all team members that other models do not demand. Deeply integrated project dependencies require problems of all types to be addressed openly and quickly. This means hard conversations must be addressed rapidly or the fiber of the approach and the team will erode very quickly. The same urgency is necessary in evaluating product owners, team members, and vendors against expectations. The method allows for adjustments and flexibility, which should be utilized to the project's advantage. For example, the product owner for one key subject area was over allocated on other strategic projects and corporate demands. As a result, management agreed to halt and postpone that segment to focus more on the remaining segments. Realization of the need for this adjustment in a more traditional method would have not been nearly as efficient and would have cost the project time, money, and resource allocation. Recognizing and addressing the issues requires diligence and focus, but the by-product for this team was trust. Agile MDM is a chaotic process and is deadline focused, but this team and leadership group worked together, kept smiling, and the endeavor was incredibly rewarding.

Core team dynamics are built through the fire of the deadlines, but those outside the core team do not have the same level of communication or constant trust development. DO NOT underestimate the impact (pace of change, process modifications) on those outside the team and how much effort this will take. Simply gathering and exposing data cannot be the focus or the determinant of success. The team must evaluate the processes that are bringing the source data to be mastered. Hindrances of data flow or integrity at the source can cripple the most well-oiled teams, and this cannot be emphasized enough. An accelerator used in this engagement to properly align the data with the process was the LEAN methodology for process and system mapping. We recommend this method for similar engagement as the approach is dependent upon the alignment of the business and data logic working together. The amount of knowledge and details that were exposed by using this model have been priceless. The people engagement that comes from the exercise itself and the ownership of the data flow starts to become reality instead of theory. There will be a lot of noise – again, keep smiling; this too is incredibly rewarding.

Divide your challenges into four root cause categories: Data, Technology, Process, and People. It takes some time to understand this; more often we want to jump into a situation and focus on that solution without fully understanding the challenge. When the root cause category is known we are better equipped to take the right stakeholders and expertise to the table. When we can communicate the true association to these categories we can have more concise discussions with all stakeholders and, by minimizing assumptions, help everyone understand. Believe it or not, this also builds trust and buy-in.

The question for this audience – How is it possible with this much focus on people and planning for this project to run successfully in an Agile manner? In reality, Agile enables the team to engage the stakeholders and continue to clarify expectations, eliminate assumptions, and check validity of the capabilities delivered often. Challenges are identified quickly, impediments are escalated quickly, and areas that need iteration are identified more rapidly. Stakeholders have embraced their stake in the progress of the project because the interactions are direct and specific, and this, along with much closer awareness, results in further trust.

This description of this ongoing project experience shows that Agile can be used to design and deploy a multi-domain MDM in an operator, but the challenges must be faced before the rewards can be gained. Effective Agile MDM does provide rapid progress, but it is much more than just loose requirements and disjointed code. Cost efficiencies are gained by eliminating long cycles between development and acceptance, but consistent cost of time from resources is not to be underestimated.

In summary, tools and capabilities from PPDM have helped keep the definitions sharp and exceptions to a minimum, which is vital in this type of engagement. Be ready to react, but be sure to listen to the problem fully before the code defines your direction. Do not be afraid to change or redirect. It is the nature of the approach and must be embraced. The rewards do not come from rapid progress; rather, they come from consistent alignment between the expectations and the outcomes. F

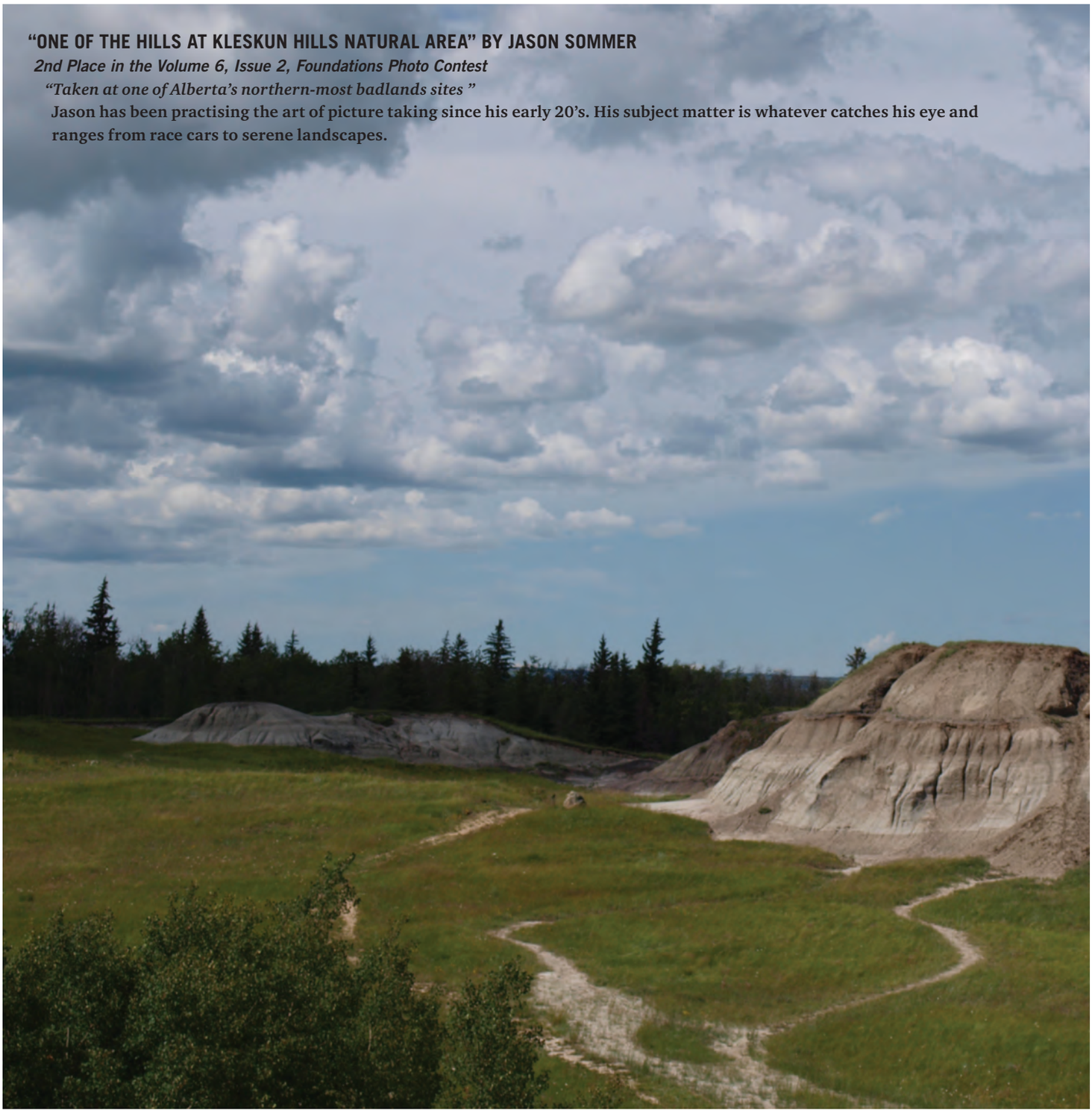
About the Authors

Melinda is the Head of Data Quality & Management for Equinor's Development & Production International business unit.

Stewart Nelson is a global leader in the Infosys Energy Practice. He is passionate about how PPDM shapes information management solutions to optimize business performance and decision making.

Photo courtesy natanaelginting / Freepik

Foundations Photo Contest



"ONE OF THE HILLS AT KLESKUN HILLS NATURAL AREA" BY JASON SOMMER

2nd Place in the Volume 6, Issue 2, Foundations Photo Contest

"Taken at one of Alberta's northern-most badlands sites "

Jason has been practising the art of picture taking since his early 20's. His subject matter is whatever catches his eye and ranges from race cars to serene landscapes.

On the cover:

"MORAINE LAKE, BANFF, ALBERTA" BY NICOLE PROSEILO

1st Place in the Volume 6, Issue 2, Foundations Photo Contest

"Alberta's beautiful Moraine Lake, hidden behind Lake Louise. Many Albertans come to its shores to hike to Larch Valley and enjoy its serenity."

Nicole Proseilo is a recent graduate and figure skating coach who loves to hike and take photos.

Enter your favourite photos online at photocontest.ppdm.org for a chance to be featured on the cover of our next issue of *Foundations*!

Feature

The Value of Data

Value and Decision Making

By John Pomeroy, The Fervid Group

That "data is an asset" is axiomatic, but few organizations have a clear idea of the true value of those assets. Oil and gas companies spend millions of dollars acquiring data and pennies maintaining it. Given that resources are finite, questions abound. Is your company investing in the quality and maturity of data types that will best serve the enterprise? Which technical data-related capabilities should we focus on – rapid ingest vs. quality profiling vs. enrichment vs. security vs. integration vs. data wrangling? Which data will be most critical to future analytics-driven use cases and what quality will be needed?

By analyzing the ways data and data processes contribute to business operations, investment decision-making, and risk mitigation, we can ensure data investments are justified, strategically aligned, and economically efficient.

SO…WHY IS IT SO DIFFICULT?

We understand well the actual value of physical assets (equipment, vehicles, buildings, etc.), but data is more problematic. Aside from the fact that data is intangible and has no physical presence, there are various factors that present challenges:

- There is a lot of data, and data is, as a rule, disorganized.
- The rate of technology churn is high and ever increasing.
- Different kinds of data and from different geographical areas are critical at different phases of the value stream (exploration, development, production, etc.).
- Data is broadly duplicated, with little lifecycle management – the same data may be in the SoR, applications, document management system, team drives, or even personal data stashes. What are the relative values of the differing instances?
- Advanced analytics presents challenges. The ability to surface unexpected relationships enables competitive advantage, but may make it harder to gauge which data may be important in advance.

WHAT PART DOES DATA PLAY IN BUSINESS?

Business is complex, and all business operations require a variety of things, to varying degrees, in order to achieve value. There are many ways to model this, noting that all activities require data and the insights derived from data.

There are two broad areas in which data provides key value:

- Maximizing return
  - Operational processes: for example, production allocation, well construction, static reservoir model development.
  - Tactical decision making: for example, well and equipment maintenance, well interventions.
  - Strategic decision making: for example, acreage acquisition, investment gates (e.g., moving from appraisal to development), well spacing.
- Minimizing risk
  - Rights assurance, reserves reporting, regulatory compliance, HSE, equity determination.

Selling data is big business – consider your personal information. The value of our personal data is well understood to the attribute level – major data brokers can sell businesses any permutation of over 10,000 attributes of personal information about each person. Unlike our personal information, oil and gas company acquired data is generally considered, for at least the initial period of their lifecycle, to be proprietary and to support competitive advantage. However, past this tight phase some data has a tangible resale value, which is sometimes seen in the sale of seismic surveys, particularly in inactive areas. Some data has a less tangible, but no less real value – for example, high-quality, well-organized data helps sell properties via data rooms.

VALUE AND DECISION MAKING

Tactical and strategic decision making in the data management world includes:

- Which data types need foundational changes; for example
  - New/enhanced SoR.
  - Data management processes.
- Which data types need data quality improvements to support current business operations and/or to support future business strategies.
- Which data technical capabilities should be implemented/enhanced; for example:
  - Rapid ingestion of large data sets to support exploration.
  - Data wrangling.
  - Data enrichment.
  - Classification/metadata extraction for unstructured data.

While challenging, it is possible to develop some methods to help inform CAPEX and OPEX data-related decision making.

As Table 1 shows, in order to make informed decisions about approach, it's important to have a good understanding of:

- Business rationale and value proposition of change opportunities.
- Degree of strategic value of change opportunities.
- Data management maturity by data domain.
- Health/maturity of technologies supporting key data management capabilities.

WHAT TO DO?

Understanding the relative criticality and maturity of data types and capabilities enables us to home in on which areas are currently fit for purpose, which areas require structural changes supported by capital investment, and which areas can be addressed by incremental changes and quality improvements.

We can assign a value proposition to each data domain in terms of how they contribute to consumers' activities, decision making, and enterprise risk at an aggregate level. Table 3 is a trivial example.

Looking at how data and data

	CAPEX	OPEX
Funding pool	Funding pool is finite and competition is fierce	Funding pool is finite
Execution	Simpler to coordinate and deliver	More complex – balancing routine operations and improvements difficult
Justification	Harder to justify major funding outlay based on poorly understood value proposition	Easier – incremental spend is lower
Organizational appetite	"Didn't we fix data management two years ago?"	Change by stealth can be easier, but changes are harder to coordinate/bundle, with concomitant risk of consumer change fatigue
Suitability for tactical outcomes	Less efficient, greater execution overhead	More efficient, less execution overhead
Suitability for strategic outcomes	Easier to execute and align platform/technology changes across the change portfolio	Technology purchase may be more difficult. Risk of duplication of effort and technology. Risk of technology conflicts/dependencies.

Table 1: CAPEX vs OPEX

	Exploration	Appraisal	Development	Production		
Data type	**Criticality**				**Data Maturity Level**	**Data Technology Health/Maturity**
Seismic	H	M	L	L	MEDIUM	HIGH
Well Header	L	M	H	H	HIGH	HIGH
Well Logs	H	M	M	M	MEDIUM	MEDIUM

Data capabilities	**Criticality**					**Data Technology Health/Maturity**
Rapid Ingest	H	M	L	L		LOW
Integration	L	M	H	M		MEDIUM
Profiling/DQ	L	H	H	H		MEDIUM

Table 2: Relative Criticality

Data Type	Cost of Consumer Data Tasks	Impact to Decision Making	Impact to Risk Management
Seismic	HIGH	HIGH	LOW
Well Header	LOW	HIGH	MEDIUM
Well Logs	HIGH	HIGH	LOW
Equipment	MEDIUM	MEDIUM	HIGH

Table 3: Value Proposition

management capabilities contribute to different business activities and phases of the value stream gives us options for identifying which components of the data management landscape will yield the best ROI.

MODELING DATA VALUE – DRIVING VALUE

There are various ways to model the contribution and value of data to various operational outcomes. These approaches include process metrics/simulation, role-based, and more complex business capability-based approaches.

Table 4 is a simple role-based example calculating the potential cost savings realized by reducing the time consumers spend on data-related tasks (column 1 in Table 3) – finding data, loading it, and managing it in applications. This model factors in consumer head count, cost/FTE, and user surveys of time spent on data-related tasks, thereby enabling calculation of data-related non-productive time. By defining a future state with per cent improvement targets in data management, it is easy to calculate savings related to specific improvements, as in the following example.

As in the example shown in Table 4, for our hypothetical organization, the annual return is high ($6.4 million) even with a modest improvement of 30% and 20% in the ability to manage and search data in apps.

By using this technique and driving down a level, analyzing the potential benefits to be realized in specific data domains, it is possible to make informed decisions on improving which data will yield the highest return.

Future work is warranted in moving to quantitative approaches from current qualitative approaches used to articulate data/information to major decision making and risk management areas.

Utilizing the best approaches in assessing data usage, understanding the role of data in your organization, and working with a consistent value model across the organization will add optics surrounding the value of data and ultimately assist in more informed and substantiated decisions at any organizational level.

About the Author

John Pomeroy is Vice President of Data Management at The Fervid Group. Over the past 35 years John has worked for software, consulting, and oil and gas operating companies.

Future State Levers	
Improvement Factor	%
Manage Data	30
Search Data	20
Transfer/Load Data	0

		Cost of Data-Related Tasks									
		Current State					Future State		Benefits		
Role	Count of Role	% Time Manage	% Time Search	% Time Xfer/ Load	Total Data NPT %	FTE MY Spent	Total Data NPT %	FTE MY Spent	FTE Savings By Role	Annual Cost/FTE	Annual Cost Savings $mm
Petrophysicist	5	30	20	10	60	3	47	2.35	0.65	500,000	0.3
Petrotech	2	20	20	40	80	1.6	70	1.4	0.2	200,000	0.0
Geotech	10	10	50	20	80	8	67	6.7	1.3	200,000	0.3
Geologist	50	10	40	10	60	30	49	24.5	5.5	600,000	3.3
Engineer	40	10	40	10	60	24	49	19.6	4.4	500,000	2.2
Data Managers	6	50	30	10	90	5.4	69	4.14	1.26	200,000	0.3
									Totals		
					FTE/year	72.0		58.7	13.3		6.4

Table 4: % Time Spent

Job Family Grid

PPDM's Professional Development Committee (PDC)

By Peggy Loyd, Cortez Resources & Margaret Barron, PPDM Association

Professional Development (PD) is essential to establishing and maintaining a standard of excellence within a professional discipline. In the PPDM context, PD addresses the needs of individual data discipline professionals, organizations that employ data discipline professionals, supervisors, and other related administrators, companies that provide data and data services, and Human Resources groups.

PPDM formed the Professional Development Committee (PDC) in late 2015. It is currently made up of 10 industry data discipline Subject Matter Experts (SME).

The remit of the PDC focuses on initiatives including the development of standards for data discipline job descriptions, career path recommendations, compensation, competencies, and a centralized repository of educational and professional development opportunities. This Committee is essential to developing and maintaining PPDM job standards and to ensuring that we are continually revising and staying abreast of the ever-evolving job roles in our companies around the globe.

The PDC engages five work streams: Job Families, Surveys, Outreach, PD Catalogue, and Value Proposition. The Job Families group, under the direct leadership of PPDM volunteer Patrick Meroney, has created a Job Families Grid (Figure 1), which is intended as an HR guideline. This is one example of the PD Committee's efforts to build a professional discipline that will solidly define petroleum data disciplines on the global career map.

The PDC has mapped the job grid to a standard set of seven sample roles. Our broader objective is to map the knowledge and skill set needed by multiple data discipline professionals, arming them with tools to develop and continuously improve knowledge and skills to help maintain high standards and engage best practices.

Through pursuit of the Job Grid initiative we are raising awareness, gaining widespread support, and enhancing credibility in the arena of data discipline professional and career development. Creation of the Job Grid initiative will enable the membership of PPDM to utilize the many training tools on our website. We foresee PPDM becoming one of the top industry organizations, with a reputation of leadership at the forefront of career and educational development.

We believe the Job Grid tool will become an industry standard, helping oil and gas companies and their employees around the globe become more educated about where they fit within their company's organizational structure. In addition, the grid will help individuals to access and educate themselves with our seminars, symposia, and online learning tools.

The Job Families toolkit is the groundwork for organizations, hiring managers, and petroleum data managers. This toolkit will include job descriptions common for petroleum data management roles. Each role will have competencies, which include skills and attributes necessary for career success.

Furthermore, we plan to create an access to career data to guide our corporate employers in making informed decisions. Within the career data structure will be the ability to review and compare salary scales and career path data to

PPDM, Professional Development Committee (PDC) Job Families Grid				
Levels				Sample Roles
Level 1 (L1)	Level 2 (L2)	Level 3 (L3)	Level 4 (L4)	(Note: Red indicates roles PDC will map competencies to)
Corporate/IT	Upstream	Strategy/Governance		Chief Information Officer, Architect
		Standards/Policy		Data Analyst, Chief Data Manager
		Projects		Business Analyst, Project Manager
		Digital Transformation		Data Scientist, Data Engineer
	Midstream/Downstream			Phase 3
Upstream	Subsurface	Geoscience	Wells	Data Manager, Geoscience Technician, Data Analyst
			Seismic	Data Manager, Data Specialist, Data Steward
		Engineering	Well (D&C)	Data Manager, Engineering Technician, Data Specialist, Data Steward
			Reservoir	Reservoir Engineering Data Manager, Reserves Data Specialist, Data Steward
			Production	Data Manager, Data Specialist, Data Steward
		Geospatial		Subsurface GIS Data Manager, Spatial Data Specialist, Data Steward
	Surface	Facilities Engineering	Engineering	Engineering Data Manager, Operations Data Specialist, Data Steward
			Geospatial	Facilities GIS Data Manager/Specialist, Data Steward
	Land			Phase 2
	Finance			Phase 2
Midstream				Phase 3

Figure 1: PPDM Job Families Grid

accelerate growth for each individual.

The training and certification for each role will be mapped to competencies and identify education and training pathways for advancement. Clear certification goals (e.g., CPDA) for each individual role will be provided.

A compensation survey was created by the Surveys work stream to help facilitate some of the detailed information to be linked to the grid. With more than 500 responses, we have gathered a significant data set to support benchmark scales and baseline information. Figures 2 – 4 are sample survey questions and responses.

Responses were received from eight of the nine global regions. Further analysis may reveal whether the number of responses from each region are representative of the number of data discipline professionals working in each region, although this may be difficult to verify.

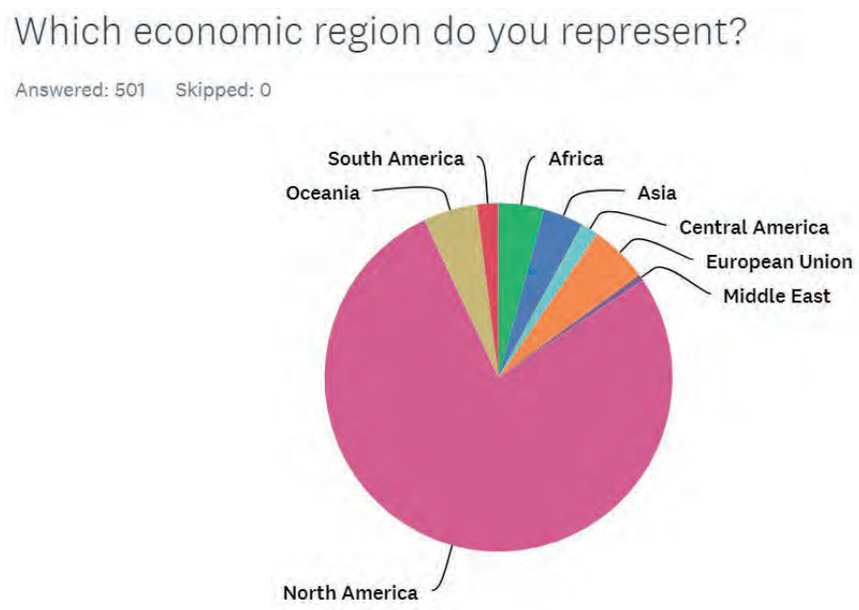


Figure 2: PPDM Compensation Survey, Question 1 Responses

Phase 2 will involve the distribution of a second survey to gather more specific job role data and is expected to be launched in the fall of 2019.

In summary, our primary goal for the Job Grid is to establish an industry standard matrix to help companies identify job descriptions and competencies within petroleum-related companies. There are many different types of oil and gas companies today: Upstream, Downstream, Midstream, and simple corporate IT-type entities. There is much to be learned in looking at organizations around the globe.

The PDC will continue to analyze and study the various roles designated within each company and will update the grid and roles to reflect these changes.

A draft Job Grid is an example of what we are working towards in creating the Job Grid Matrix.

If you would like to get involved as a volunteer with the Professional Development Committee or one of the work streams, please contact volunteer@ppdm.org.

PPDM would like to thank the following members of the PDC for their ongoing dedication to this body of work:

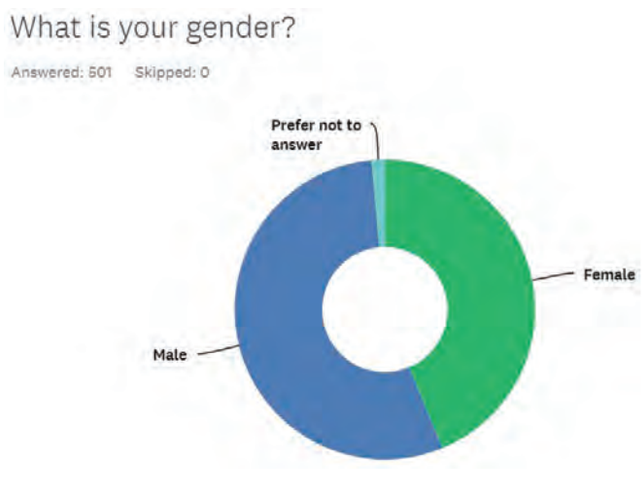


Figure 3: PPDM Compensation Survey, Question 2 Responses

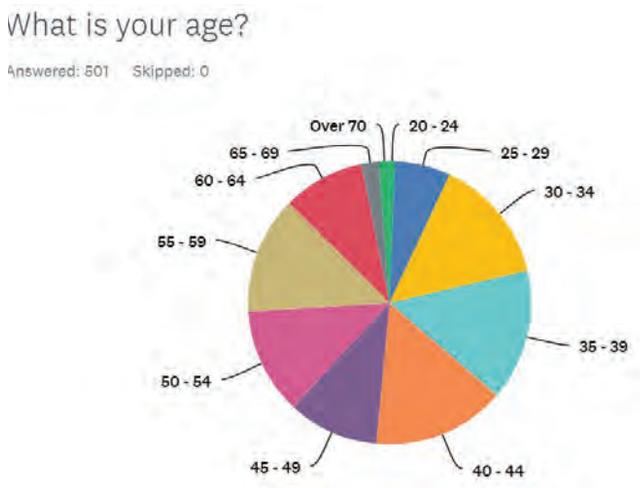


Figure 4: PPDM Compensation Survey, Question 2 Responses

PDC MEMBERS

Cynthia Schwendeman, Chair, Team Lead, Subsurface Information Management at BP

Pat Meroney, Job Families Work Stream Lead, Vice President US Operations and Professional Services, Katalyst Data Management, USA

Oliver Thistleton, Senior Consultant / Operations Manager, Asia Pacific, DataCo, AUS

Dean Melo, University of Aberdeen, formerly Petrobras, UK

Maria Carandang, Surveys Work Stream Lead, Data Analyst, Shell, USA

Tracy Heim, Information Management Specialist, AER, CAN

Tricia Ruud, Senior Geological Technician, Warwick Energy Group, USA

Peggy Loyd, Outreach Work Stream Lead, Cortez Resources, LLC, USA

Cindy Cummings, Exploration Digital and Physical Data Coordinator, Repsol, USA

Zubai Abu Baker, Exploration Database Analyst, Repsol, MYS

Margaret Barron, Chief, Professional Development, PPDM Association

About the Authors

Peggy Loyd, Vice President at Cortez Resources, 30+ years of oil and gas experience. As a PPDM volunteer, she serves as a member of the PDC and is the lead for Community Outreach and a member of the DFW Regional Leadership Team.

Margaret Barron has worked in education and training for over 25 years. Her work with the PPDM is focused on advancing the professional discipline for petroleum data managers.

A PPDM Vision For Business Rules

By Dave Fisher, David Fisher Consulting

Good decisions require good data. Rules-based quality assurance builds trust in the data.

Would you trust your bank statement if it was handwritten by the receptionist? Would you trust your child's report card if she wrote it herself? Would you trust your car repair to a bicycle mechanic? Do you believe everything on social media?

Quality assurance is *the part of quality management focused on providing confidence that quality requirements will be fulfilled*. ISO 9000.2015 https://asq.org/quality-resources/quality-assurance-vs-control.

Quality assurance (QA) is an essential component of data governance. It aims to prevent errors from occurring in the first place. This is surely much better than detection, tracing, repair, and recovery.

THE QA VISION

The PPDM Association's vision is that QA will be an integral part of data management from creation to exchange to decision-making:

- Semantics: sender and receiver understand the same meaning in the data name/label.
- Expectations: sender and receiver have the same basis for trusting the data.
- Processes: industry-agreed criteria and procedures to create / store / retrieve / use the data.
- Validation: assurance that the data can be trusted by everyone who handles and uses it.

If we achieve this vision, what would the new world look like?

- Done right first time.
  - Regulatory filing.
  - Partner data exchange.
  - M&A data merge.
  - Corporate data is "golden" (trusted by all).
- Save time, money, and opportunity.
  - Real-time access to trusted data means decisions are faster and better.
  - Data managers focus on real value, not repair.
  - Avoid risk and reworked business plans caused by bad data.
- Enhance and extract the value of data assets.
  - You paid for it, now unlock the value (ROI).
  - The greatest value is generated by using the information to make successful decisions.

MEASURING QA SUCCESS

The Capability Maturity Model (CMM) was devised to improve software development processes. Adapting the CMM to data QA, we can say that an organization (or ideally, an entire industry) achieves CMM Level 4 maturity when:

- Data QA process objectives can be evidenced across a range of operational conditions.
- The suitability of the QA process in multiple environments has been tested and the process refined and adapted.
- QA users have experienced the process in multiple and varied conditions and are able to demonstrate competence.
- The QA process is adapted to projects without measurable loss of quality or deviation from specifications.
- Process capability is established and measured. Example: data transfer and integration from Party A to Party B is done 100 times with zero negative outcomes.

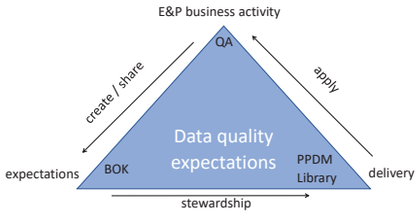


Figure 1: Data Quality Expectations

There's a long road ahead to this level of QA! The journey will only be successful if we work together.

PPDM RULES LIBRARY

PPDM has a library of data rules (rules.ppdm.org.) A data rule is *a statement that provides the opportunity to validate compliance to expected conditions*. It is an atomic test (one data item at a time) that resolves to true or false. The intention has always been to receive rules from our members. To date, only a few members have submitted some rules. Others have expressed support for the idea of a library, but have declined to contribute because their rules are proprietary.

Everyone is a stakeholder in industry data. Although we all create and manage data internally, no organization is an island; we need external sources. How you manage data internally is your own affair. But as soon as you share or report this data to a third party, your data quality can affect the

entire industry. If you do not "vaccinate" your data, you may cause an epidemic.

As with the fight against infectious diseases, prevention must apply across the entire industry.

> The best place for QA is in the "home" (where it originates), not in the "hospital" (treating the disease).

Some say, "Our data rules are proprietary." Yes, your company invested time and money to create these rules. But is that your competitive advantage? Perhaps so, if you only consider your internal data. But as soon as you import data from a partner or other third party, you incur the risk of "infection." Your advantage is not in your rules; it's in your data governance and how you use the data.

The industry uses standards and specifications from IOGP, API, etc., for positional surveys, tubular goods, gas analysis, etc. Why insist on your own secret QA rules and methods for data?

BUILDING TRUSTED DATA

Do you trust your own data?

The way you create and manage your data will eventually affect all of us.

- Do you ensure trusted data at the point of creation?
  - Are agreed industry expectations built in to the way the data is generated?
  - Is the data fit for future uses or only for the purpose at hand?
  - Are QA procedures applied before the data is accepted from the system or contractor?
- Do you know and measure the cost of QA?
  - Creation: QA at this stage is an investment in quality (trust) through the life of the data and the business asset.
  - Tracing: how far does a data error travel through your internal systems and decisions before it is identified and fixed?
  - Fixing: what happens if a fault is discovered after delivery?
    - Internal business units, databases, and applications.
    - External partners, customers, regulators: are they notified?
  - Risk: how does your bad data impact other parties and how may it rebound on you?
- QA and the data life cycle
  - Do you QA the same data more than once? Why or why not?
  - Do you QA at source (e.g., well site or desktop) or at destination (e.g., desktop)?
- Combining data from multiple sources is a race to the bottom unless the same QA is applied.

CAN YOU TRUST EXTERNAL DATA?

External data is data from partners, suppliers, data vendors, and governments.

- How do you define and measure "trust?"
- Are your data quality expectations embedded in contracts?
  - How is contract performance measured?
  - You pay more for "trusted data," but how do you measure the added value?
- Do you apply the same QA rules to external sources?
  - Before transformation – your coded rules probably don't work.
  - After transformation – semantic errors are already in place.
  - Merging internal and external data – what happens to quality?
- Is your trust hierarchy based on source, not on QA?
  - Do you consider that Source A is always better than Source B?
  - If you always put your own data at the top, you may be missing something better.
- Do you have assurance that the external party will send corrections when discovered?

CAN OTHER PARTIES TRUST YOUR DATA?

You send your data to partners, governments, etc.

- How do you ensure compliance?
- How do you fix errors that are discovered after delivery?
- How do you measure the risk of delivering bad (wrong, incomplete, non-compliant) data?
- Does your executive level understand the cost and risk of data submission and exchange?
- Can you justify the cost of "delivering it right – first time, on time?"

INDUSTRY INITIATIVE FOR QUALITY ASSURANCE

The E&P industry should collaborate toward the PPDM vision of Quality Assurance:

- Develop and use the same sets of QA rules.
- Compete on data governance and analysis, not on basic quality.
- The PPDM Rules Library is part of the Body of Knowledge for data management.

Incremental success is possible by working together:

- One step at a time; one success at a time.
- Share your rules to build the PPDM Rules Library.
- Help to edit and validate the rules in the Library.
- Invest in the Library – PPDM needs project funding.
- Pick a type of data that we all use; agree on some data rules.
- Agree on a business rule – the process for applying a set of data rules, such as QA rules to validate LAS well log header data or a monthly well production report.

CONCLUSION

Rules-based quality assurance builds trust in the data. Let's vaccinate our data so everyone benefits.

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.

My Adventures in Data Science

By Francisco Sanchez, Houston Energy Data Science

Through my early adoption of data science in energy back in 2011 and my affiliation with Houston Energy Data Science, I have been lucky enough to have a first-row seat to the evolution of data science in the local energy industry and the impact it has made in the past few years. I have met some of the most interesting people in the world and Houston's brightest minds in data. My journey is ongoing and growing, but I want to share just a few of the stories and events that have shaped data science in energy.

BACK IN THE DAY

I first heard about data science in the winter of 2011. At that time, I wondered what everyone in the energy industry was doing in data science. To my dismay, I found no one with this title working for energy companies, at least based on a search of LinkedIn titles. Oh, I am sure people will say that they were doing algorithms, regression, etc. at that particular time – I do not dispute that – but they were not doing actual data science. I am talking about real data science, which I define as the process of gathering empirical evidence, transforming data, documenting these transformations, testing several algorithms and models, picking the best model based on cross-referencing, and then deploying the model based on both empirical evidence and business decisions.

In this darkness, there was one shining light: Judson Jacobs' 2010 presentation at CERAWeek, where he described how the use of analytics could help improve efficiencies in operations of oil and gas assets. Mr. Judson outlined how other industries, such as railroads, had adopted predictive analytics for preventive maintenance and his thought that the energy industry should adopt such techniques to prevent failures on the oil patch.

In 2012…not much had changed.

KAGGLE GETS IN THE MIX

In 2013 I received an unexpected email from Anthony Goldstein, a member of my LinkedIn analytics group. Anthony was the CEO of Kaggle, a new startup for machine learning competitions (in later years, it was bought by Google). He wanted to see if we could meet for dinner while he was in town to meet potential clients to talk about data science in energy. Our discussion shared one common theme: disrupting the energy industry with data science. His company was already doing this, and he had procured contracts with oil and gas companies. I was pleasantly surprised at the results, which were able to more accurately predict expected ultimate recovery numbers and production numbers using a technique called Ensemble Methods. These methods were so successful at prediction that Kaggle was contemplating building a platform for use in future consulting gigs.

Kaggle was the first data science consulting company to concentrate their efforts in the energy industry and successfully deploy predictive analytics models. The good times did not last long; as we all know, the market tanked and so did the contracts for Kaggle. So, Kaggle decided to leave oil and gas and concentrate on competitions, but not before making a huge impact on data science in energy.

TERR AND THE START OF HEDS

In the summer of 2014, TIBCO invited me to attend a Meetup event they were sponsoring. Their chief data scientist, Michael O'Connell, was to present a new product in Spotfire, their already-successful Data Viz software. This new product was called TERR. It was free food and wine, so why not! I was also interested in TERR because it was the first time I had seen R written in a Data Viz product and regression code run outside of R Studio. The Meetup was run by the R users group. It was the first time I had seen so many data people together in Houston; granted, there were very few data scientists in the group. This Meetup gave me the idea of opening a Meetup of my own, one that would bond data science / data people with energy professionals. Thus, Houston Energy Data Science (HEDS) was started.

In November of 2014, HEDS held its first Meetup with a whopping 10 people. The Texas A&M MS in Analytics program got wind of my Meetup and wanted to become part of the movement. At the time, the MS in Analytics was the only program in Houston offering data science

Photo courtesy Jannoon028 / Freepik

training, and it was in its infancy. Today, HEDS brings between 60 and 120+ people to our Meetups and has been sponsored by many industry companies. HEDS has helped energy industry professionals understand the potential of data science while introducing many data scientists to the exciting world of the energy industry.

EARLY DATA SCIENCE TEAMS TAKE FORM

By 2014, a few energy companies had taken the plunge into data science and were starting to form what we would call Data Science departments. Most notable were Chevron, ConocoPhillips, and Devon. In services, Schlumberger took the lead when it started to hire data scientists in late 2014 to early 2015. Hands down, the leader of the pack was Devon.

Devon had bought all the bells and whistles that SAS had to offer and built a strong analytics department, almost entirely organically, with very strong training and full support by upper management. Some of the items they were starting to predict were remarkable, but their most remarkable achievement in the early days was the integration of all of their data silos into one. The connections made, and insight found, just by putting several types of data systems together made all the difference in creating predictive models. Devon is still a leader in the use of data science, but majors such as ExxonMobil, Shell, and Chevron are pushing through data science initiatives.

THE NEED FOR SPEED

At the beginning of 2016 and into 2017, as data science teams in energy companies were becoming more accustomed to energy data and the industry (for PhDs not used to this data), certain problems started to pop up. Energy companies had a vast amount of data, and it was difficult to work with this data, which lived in several database silos. Hortonworks and Cloudera had a great solution in Hadoop, which a lot of energy companies loved. Not to get too technical, Hadoop allowed the distribution and processing of large amounts of data into several nodes while making copies of this data so as to mitigate failures if a node was to be down. Hadoop would allow you to put large amounts of data together, like ingredients in a sandwich, at amazing speeds. Eventually, Spark would become the darling for processing large amounts of data, reducing the number of read/write operations to disk.

AUTOMATION SENSATION

Going into 2018 and today, automation became the biggest push for energy companies. How could we process data fast and make sense of it without having to do it in code? Sounds crazy, right? A lot of data science work in energy cannot be automated because much of the most important work is in choosing a business strategy or business solution based on the data. Other tasks, however, can be automated. If we want to combine data and munge it without using SQL or another language like Python or R, options like PowerBI/Power Query and Qlik Sense makes this easy to do on their platforms. Some companies allow you to do some of the data science work without having to write code; two popular ones are Alteryx and Rapid Miner. If you are using supervised methods of predictive modeling, DataRobot will allow you to cross-reference your models and will help you choose which model works best on your test data.

THE FUTURE IS HERE!

In the future, data scientists in energy will need to learn how to create products with data, such as artificial intelligence models that learn and process data and virtual reality environments, while also learning the new way to transfer data in the Blockchain.

By no means is this a full taxonomy of data science in energy; there are a lot more data stories to be told, including more events that have shaped this field. My hope was to briefly present some very important events that changed data science in energy here in Houston.

About the Author

Mr. Sanchez has over 20+ years as a professional in Accounting, Finance, Data Science, and Business Consulting.

Thanks to our Volunteers

MAY 2019

Cora Poche

Congratulations and thank you to Cora Poche, our May 2019 Volunteer of the Month. Cora was a volunteer on the Houston Leadership Team for several years, acting as Chair during the challenging transition from the Houston Symposium to the now highly successful Houston Expo. "Cora has been a tremendous asset to PPDM and through her work on the Houston Leadership Team, she has provided and been regarded as a source of advice, support, and more for us. Cora truly deserves this recognition," said Pam Koscinski, USA Representative with PPDM.

Cora is a Principal Technical Data Management Advisor at Shell. She has spent the majority of her career in technical and leadership positions covering Subsurface Well Data Management (DM), Seismic DM, Global Rock and Fluids Properties DM, Global WRFM DM, Petrophysics DM, and "TDM in the Business." In her current advisory role, she leads multiple data and information management projects across the globe. Cora chaired the Houston PPDM Leadership Team in 2017-18. She also served on the Advisory Board for the PNEC Data Management Conferences and was Advisory Board Chair in 2015. Cora received the PNEC Cornerstone Award in 2018 for significant data management contributions to the industry. She is a member of the Houston Geological Society.

JUNE 2019

Patrick Meroney

Patrick Meroney was PPDM's June 2019 Volunteer of the Month. Patrick has been a member of the Professional Development Committee since 2017. "Patrick has been an essential part of our Professional Development Committee. He also works with the Katalyst Data Management team to make so many of our events possible. We are fortunate to have an ambassador in Patrick," said Margaret Barron, Chief, Professional Development with PPDM.

Pat Meroney started his journey in Data Management working for Conoco in 1990, holding various positions at Conoco and ConocoPhillips in and out of Business Roles and IT, but primarily focused on Data Management, Technical Computing, and Strategy. In 2011, he joined the Spanish company Repsol, where he was tasked with building out the Information Management group in Houston to support the North American Business and Western Hemisphere Exploration teams. Pat joined Katalyst Data Management in 2015, and now serves as Head of US Operations. He lives in colorful Colorado.

JULY 2019

Grace Yang

Secretary of the Australia West Leadership Team, Grace Yang was the July 2019 Volunteer of the Month. "Grace has been a tremendous help with the activities going on in Perth and the surrounding area for PPDM. Frequently quiet, behind the scenes organizing events, Grace has been instrumental in handling many onsite logistics and keeping PPDM events in Perth possible," said Elise Sommer, PPDM's Senior Community Development Coordinator. "It is through volunteers like Grace that we are able to bring the data management community together around the globe."

Grace Yang is the Director of Finance, Asia Pacific, for Katalyst Data Management. Prior to this new role, she was the Finance Manager at Katalyst, and Finance and Administration Manager for Spectrum Data. Grace also spent several years with AXS Access Management, De Jong Hoists/ Rimwest Pty Ltd and IFS Construction. Grace is a Chartered Accountant and has a Master of Business Administration and a Master of Professional Accounting from the University of Western Australia.

SEPTEMBER 2019

Cynthia Schwendeman

Congratulations and thank you to Cynthia Schwendeman, our September 2019 Volunteer of the Month. Cynthia is the chair of the Professional Development Committee (PDC) and has brought energy and enthusiasm to this team with a big remit. "Cynthia's dynamic energy and commitment to the work of the PDC have profoundly impacted the tempo and subsequent outputs of this highly valued group of volunteers," said Margaret Barron, Chief, Professional Development. "Under Cynthia's leadership, the PDC is on track to meet goals set for 2019 – 2020. We're very grateful to have such a passionate leader at the helm of the committee."

Cynthia is the Team Lead, Subsurface Information Management at BP. Cynthia has more than 16 years of experience in E&P subsurface data management and is considered an expert regarding well data quality in geological interpretation software. As a Data Advisor for Hess Corporation, Cynthia was responsible for a program of work that created systems and processes to proactively deliver geologic data to the Bakken asset. In her role at Occidental, she was responsible for geological data in the Permian Enhanced Oil Recovery team. Prior to joining Occidental in 2018, Cynthia spent approximately 10 years at Hess Corporation. Cynthia holds a bachelor's degree in Geology from Louisiana State University.

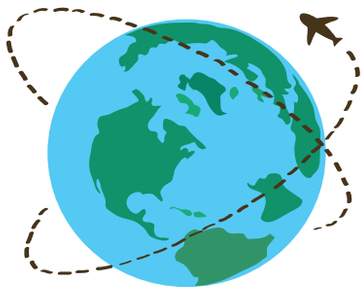
Brisbane Data Management Mini-Workshop 2019

Adelaide Data Management Luncheon 2019

Perth Data Management Workshop 2019

Oklahoma City Petroleum Data Workshop 2019

PPDM Around the World



Houston Professional Petroleum Data Expo 2019 - Board Panel

Houston Professional Petroleum Data Expo 2019 - Keynote Speaker

Dallas/Fort Worth Petroleum Data Workshop 2019

Calgary Data Management Symposium, Tradeshow & AGM 2018

Denver Petroleum Data Symposium 2018

UK Data Management Luncheon 2018

LUNCHEONS

NOVEMBER 5, 2019
OKLAHOMA CITY DATA MANAGEMENT LUNCHEON
Oklahoma City, OK, USA

NOVEMBER 7, 2019
MIDLAND DATA MANAGEMENT LUNCHEON
Midland, TX, USA

NOVEMBER 2019
PERTH DATA MANAGEMENT LUNCHEON
Perth, WA, Australia

NOVEMBER 2019
BRISBANE DATA MANAGEMENT LUNCHEON
Brisbane, QLD, Australia

DECEMBER 3, 2019
TULSA DATA MANAGEMENT LUNCHEON
Tulsa, OK, USA

DECEMBER 10, 2019
DALLAS/FORT WORTH DATA MANAGEMENT LUNCHEON
Dallas/Fort Worth, TX, USA

JANUARY 14, 2020
HOUSTON DATA MANAGEMENT LUNCHEON
Houston, TX, USA

FEBRUARY 5, 2020
LONDON DATA MANAGEMENT LUNCHEON
London, UK

FEBRUARY 5, 2020
CALGARY DATA MANAGEMENT LUNCHEON
Calgary, AB, Canada

FEBRUARY 11, 2020
DENVER DATA MANAGEMENT LUNCHEON
Denver, CO, USA

FEBRUARY 19, 2020
BRISBANE DATA MANAGEMENT LUNCHEON
Brisbane, QLD, Australia

FEBRUARY 20, 2020
OKLAHOMA CITY DATA MANAGEMENT LUNCHEON
Oklahoma City, OK, USA

WORKSHOPS, SYMPOSIA, & EXPOS

OCTOBER 21 – 23, 2019
CALGARY DATA MANAGEMENT SYMPOSIUM, TRADESHOW, & AGM
Calgary, AB, Canada

NOVEMBER 13, 2019
DENVER PETROLEUM DATA SYMPOSIUM
Denver, CO, USA

FEBRUARY 25, 2020
DALLAS/FORT WORTH PETROLEUM DATA WORKSHOP
Irving, TX, USA

MARCH 31 – APRIL 1, 2020
HOUSTON PROFESSIONAL PETROLEUM DATA EXPO
Houston, TX, USA

MAY 12, 2020
OKLAHOMA CITY PETROLEUM DATA WORKSHOP
Oklahoma City, OK, USA

CERTIFICATION - CERTIFIED PETROLEUM DATA ANALYST

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

MAY 13, 2020
CPDA EXAM
(Application Deadline April 1, 2020)

NOVEMBER 4, 2020
CPDA EXAM
(Application Deadline September 23, 2020)

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 2020. Public training classes are also planned for the rest of 2019 and 2020.

OCTOBER 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		

NOVEMBER 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

DECEMBER 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				

JANUARY 2020

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 2020

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

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