Final Report

submitted to the Austrian Marshall Plan Foundation

Petroleum Data Analytics

Clemens Rainer
Montanuniversität Leoben, Austria

Abstract

This research focuses on the present state and future needs of digital skills for Petroleum Engineering students. The first section presents current understandings of why and how curricula should be adapted to keep academia in line with the needs of the industry and emerging technologies. Next, an overview is given of the specific petroleum engineering and data science disciplines. The chapter *Digital Transformation* portrays how digitalization has shaped the oil and gas industry, its current and future use cases. An explanation is given why data analytics is key to manage the increasing availability and amount of data for engineering and business purposes.

Two graduate-level data analytics courses, specific for petroleum engineering, are built as online coursework and outlined together with Colorado School of Mines particulars. This report concludes with examples of how to analyze and interact with oil and gas specific datasets.

Keywords: petroleum engineering; data analytics; graduate curriculum; online course; reservoir engineering; drilling engineering; production engineering; automation; digital transformation; digital oilfield; data science; Industry 4.0.

Acknowledgment

This project was funded by the Austrian Marshall Plan Foundation. *I would like to thank my supervisors Prof. Alfred Eustes at Colorado School of Mines, and Prof. Herbert Hofstätter at Montanuniversität Leoben.*
Further thanks go to the Department Petroleum Engineering at Colorado School of Mines, Golden CO, for hosting my stay, especially Jim Crompton for the original course content and the Trefny Center for coordinating the course development process.

Contents

---

1. Introduction

A current skillset for efficient processing, filtering, analytics, interpretation and attention to relevant information is needed for forthcoming engineers in a digital environment. The round-the-clock access to data and its increasing volume changes how we work together and have access to a wealth of resources.

Petroleum Engineers (PE) deal with a wide variety of data on a day-to-day basis, software tools and programming skills allow them to center on creating solutions and making informed decisions. The information technology (IT) landscape changed greatly in recent decades and a new generation of smart technologies became available for everybody, such as open-source software, public machine learning (ML) frameworks and research, online collaboration and sophisticated technologies as a service. In the oil and gas industry commercial technologies are prevalent, however, being able to interface different products allows engineers to come up with novel solutions and reduce time spent on manual data processing and labor-intensive manipulations. Furthermore, the next generation of engineers needs to be attracted to working in the energy sector, as trending industries and the high tech sector hire the best talents with IT and data skillsets.

State of the art logging-while-drilling tools communicate real-time sensor readings at a slow speed of a few bits per second, whereas in seismic acquisition terabytes of data are generated within minutes. This scale of available information is true for all sub-disciplines from geophysics, reservoir, production, subsurface, wells engineering, economics and as such ubiquitous across the vertical value chain of a fully integrated oil company or service provider.

The amount of data recorded and processed is growing exponentially and requires current methods and skills to efficiently extract knowledge and come up with fresh solutions. The quality, frequency, format, and accessibility varies for the various areas of petroleum engineering. For example, in mature oil and gas assets, well pressure data is read off manually from gauges every few days whereas in new installations downhole pressure gauges report real-time information continuously. Engineers may need to process and combine different qualities of data, even going back to analog readings recorded decades ago, in the same analysis.

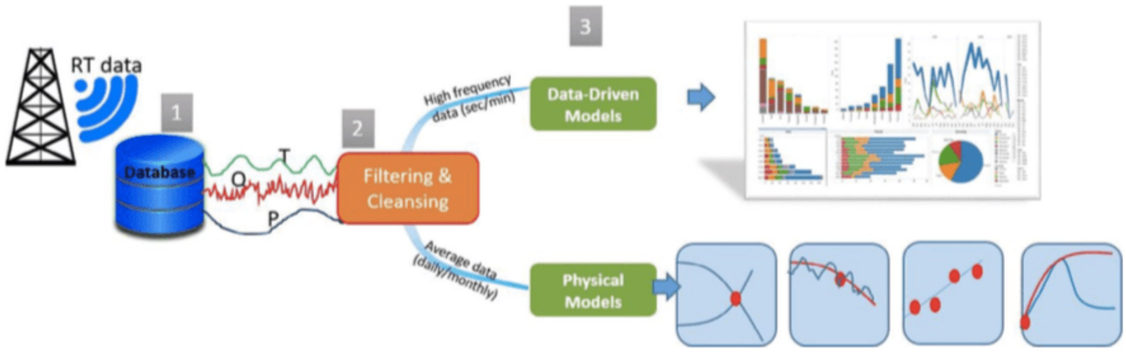

Figure 1: Example of an automated workflow in the Digital Oilfield. Real-time data, such as pressure p, temperature T and flow rates q are transmitted and stored in a time-series database (historian). Data analytics is performed in steps 2 and 3, either as source for physical models or bulk processing in data-driven analytics [1].

For this, an understanding of data preparation, processing, analytics, and respective shortcomings is needed to make informed decisions and prevent errors from being introduced. A basic IT skillset helps to cope with the growing amount of available data and interface efficiently with the wide variety of (proprietary) hardware and software tools used in the industry.

Recent advances in data science and processing open a wide field of applications and the pace of technology change is rapid. As the complexity of operations and amount of data grows by orders of magnitude the next generation of engineers and decision-makers need a good basic understanding of data analysis methods combined with practical experience and domain knowledge.

General IT knowledge, such as computer networks, retrieving information online and interfacing various operating systems require little attention as the technologies became mainstream. However, a comprehension of how to interact with proprietary tools and open source software is lacking. This includes basics in computer programming and interfaces between hardware and software. An example is being able to process and prepare raw data across disciplines for specific applications to derive optimal solutions. Different levels of old and current technologies need to be aligned. This can be done with programming and data processing. These disciplines are rapidly evolving and gaining momentum. Basic skills in programming, statistics and analytics are required to succeed in a digital environment and benefit from technologies such as machine learning, pattern recognition, IIoT (Industrial Internet of Things), digital twins, cloud and edge computing.

2. Motivation

The 4th industrial revolution will change how work is done and utilizes automation and processing to help engineers focus on making informed decisions with less attention to manual data input, manipulation, and analog workflows.

This is especially true to the oil and gas industry with its gradual adaption of technologies as conventional field development projects span decades, historic high-profit margins, an aging workforce and a focus on safe and proven technologies.

Technologies of real-time communication and processing big amounts of data became mainstream and most industries are currently changing due to the "digital revolution" (see figure 1 as example). The oil and gas industry is cautious to adapt new and upcoming trends and the same is true for the general Petroleum Engineering (PE) curriculum that has seen only minor updates during the past decades.

The requirements, skillset, and expectations for the next generations of petroleum engineers need adaption to address the needs of a changing industry. The most recent *Colloquium on Petroleum Engineering Education* formulated the following outline regarding the future of the petroleum engineering academic training [2]:

The "century-old" petroleum engineering curriculum is currently under revision to make it fit new technologies and attract the next generation of engineers.

1. *Review of the century-old curriculum; Changes in the industry; Changes needed to bring the curriculum up-to-date with industry's expectations.*
2. *Incorporation of Petroleum Informatics in the curriculum.*
3. *Developing a new sense of identity for PE graduates.*
4. *What is obsolete? What additional skills should we incorporate in the education of PE graduates to make them marketable to other related industries?*
5. *Should the emphasis areas be changed to incorporate more courses on data driven analytics and the concepts of artificial intelligence?*
6. *Considering application of the PE concepts to other subsurface resources, is it appropriate to emphasize subsurface engineering as an identity for the PE graduates?*
7. *Industry is moving towards automation. Should there be curricular changes in PE to incorporate these advances in IOT's?*
8. *What are the best practices in faculty recruiting and retention; diversity; public relations and outreach?*
9. *What are the issues and solutions in funding from government and industry, and what are the implications?*
10. *What are the sustainability issues? Should a PE degree become a professional degree with a 5-year education program ending with a MS degree? How about the issues related to proposed changes when it comes to global PE education?*

This research project specifically addresses points **1, 2 and 7** from the outline presented above.

Motivation regarding the present-day relevance for expanding data analytics and programming skills lectures for petroleum engineers at the Colorado School of Mines [3]:

> *"The next generation oil and gas engineering workforce must be prepared for this new reality and have the skills to adapt to and thrive in a more data-driven work process and an interdisciplinary collaboration work environment. CSM needs to be in the forefront of the education of the petroleum engineering workforce. We have started that process with the undergraduates. We now need to expand this effort to the graduate level and beyond."*

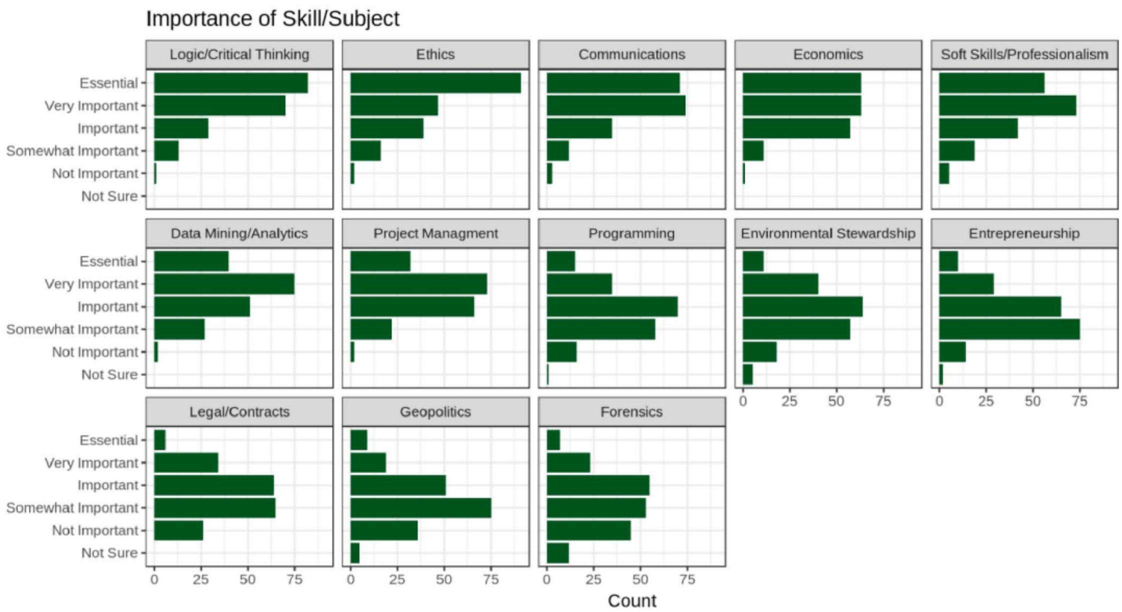

Figure 2: Normalized results of an 2019 study about important skills for new PE graduates ranked by importance [4].

3. Related Work

3.1. PE Education

There have been numerous studies and papers published about the state of PE education, competencies and training for the oil and gas industry. The Society of Petroleum Engineers (SPE) infrequently hosts a colloquium on the state of petroleum engineering education and addresses changes and additions to the curriculum for PE degrees. The goal is to align industry needs for the future workforce with academia by setting R&D goals, knowledge transfer, partnerships and competency standards [5]. The most recent colloquium was held in August 2018 [2]. SPE maintains a competency matrix to define the skills and knowledge for the different petroleum engineering domains that serve as a reference for recent graduates and the minimum expected competencies for entry into the industry [6]. Academia and industry typically refer to the U.S. in this context [5]. Likewise, the *U.K. and Ireland Engineering Education Research Network* held a colloquium in 2018 focusing on topics alike and the collaboration between academia and industry from a European perspective [7]. Tymkiv compares the PE degree programs offered in Europe [8]. In 2019 there were twenty U.S. based and eleven international PE programs accredited by ABET (Accreditation Board for Engineering and Technology) [9].

These papers have in common that they were published after the sharp oil price drop in 2014 which forced the petroleum industry to reorganize and rethink how to do business in a more competitive, low price environment. One solution to complete more work with less staff is by applying digitalization and analytics to optimize processes and workflows. In addition to the cyclic nature of the industry and the ongoing rejuvenation of the workforce (the "great crew change"), academia has to prepare its graduates for lifelong learning and digital technologies [10].

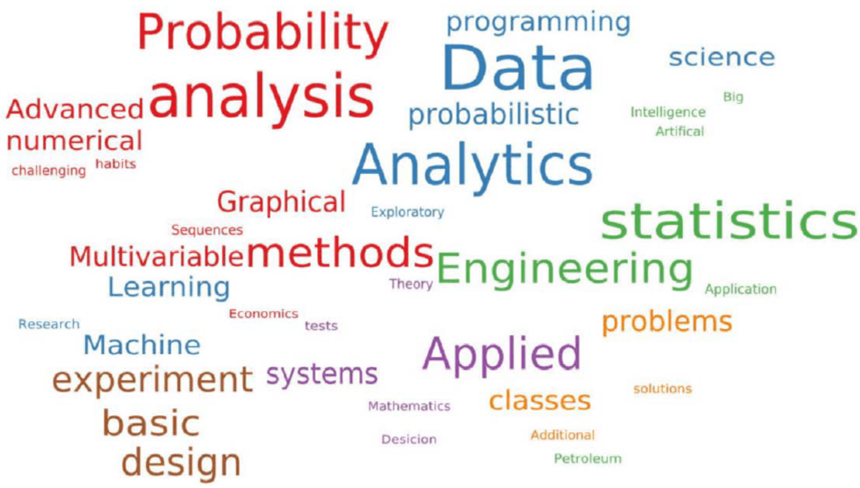

Figure 3: Word cloud highlighting skillset and tools for digital adept Petroleum Engineers. The font size is proportional to its present importance (number of occurrence) in a free response question [4].

A comprehensive survey among young industry professionals was presented in 2019 [4]. Objectives were to understand how well academia prepares its graduates for a career in the current professional environment and what skills are required and might be missing in current courses and PE programs. There is a bias because the survey looks at the U.S. industry only, where most students join the workforce with a BS degree.

Horizontal drilling technologies and hydraulic fracturing changed the North American oil and gas industry considerably, which is evident in current industry needs. Still, 75% of the responses were from people with more than 20 years of industry experience and gives a valid impression on how the industry has changed during the past decades and how well present-day graduates are prepared for a Petroleum Engineering career nowadays. Figure 3 visualizes the answers to open-ended questions asking what courses should be included in the curriculum. Common themes were emerging technologies, unconventionals and hands-on analysis of big data and real-world data sets. Furthermore, decision making under uncertainty, management of risk and knowledge of oil and gas business aspects are seen as vitaly important. One conclusion of the survey is that the industry is satisfied with the theory being taught, but teaching practical applications of PE principles is lacking. Further concerns are how easily PE graduates can transition to industries other than oil and gas and how to modernize course materials to include applications of new technologies such as data science and the digital revolution within oil and gas.

The same survey results are the basis for an article dealing with the question if four years of college are enough to prepare students for an increasingly complex PE profession and how well business aspects, sustainability, and information technologies are addressed among the different U.S. PE schools and colleges [11].

3.2. Emerging Technologies

Mathieson et al. compare the current state of Petroleum Engineering to the major engineering disciplines taught and raises the questions if PE can be recommended as a career path and will still be relevant in the coming decades. The conclusion is that PE is here to stay as fossil fuels will remain the prime energy source for the foreseeable future. However, it is suggested that current education programs need to be revisited to remain attractive so that graduates will not be locked into only one industry with characteristic cyclical downturns and massive layoffs. Equally to other STEM (science, technology, engineering and mathematics) disciplines, PE has to adapt to a future where engineers are skilled in advanced technologies and become "digital citizens". New frontiers and technologies won't replace engineers, but augment analysis and subsequent decision making on complex, real-time data to improve processes and interact with remote physical assets.

As result graduate education will become more important, as it is within exising scientific disciplines. Advanced and addtional training will not fit into an undergraduate degree for the level of sophistication required to meet the challenges the industry is phasing [10].

Ershaghi et al. present a historical look at PE education, current research in addition to applications of new and upcoming technologies. Oil and gas were among of the first industries employing complex telecommunication systems, large data processing, and numerical simulations as soon as computers became available to implement novel concepts, such as 3D visualization and modeling.

At the same time, physical extraction methods and drilling technologies evolved to allow ultra-deepwater developments, enhanced oil recovery and reservoir stimulation to increase productivity from existing fields and extract hydrocarbons from low permeable shale formations. Data informatics and automation are expected to keep oil and gas competitive in the future by increasing efficiencies, cutting costs and increasing reliability and safety. One focus will be education beyond an undergraduate degree and responding to the current and future needs of an evolving industry. Information technology and data analysis are key to identify trends and derive plausible solutions to help engineers collaborate and make timely decisions.

> *"The oil and gas industry is complex and is here to stay in the foreseeable future if it can focus on increasing its efficiency and environmental safety practices. In this transformation, it needs to import solutions from many industries including IT. To equip the professional MS level university graduates joining the industry, there is a need to incorporate new concepts related to data and event management at the graduate level."* [12]

The term "petroleum informatics" is defined and key technologies identified to be included in the curriculum [12]:

- Big data, data science and analytics.
- Machine learning and artificial intelligence (AI).
- Industrial internet of things (IIoT).
- Cyber-physical systems and security.

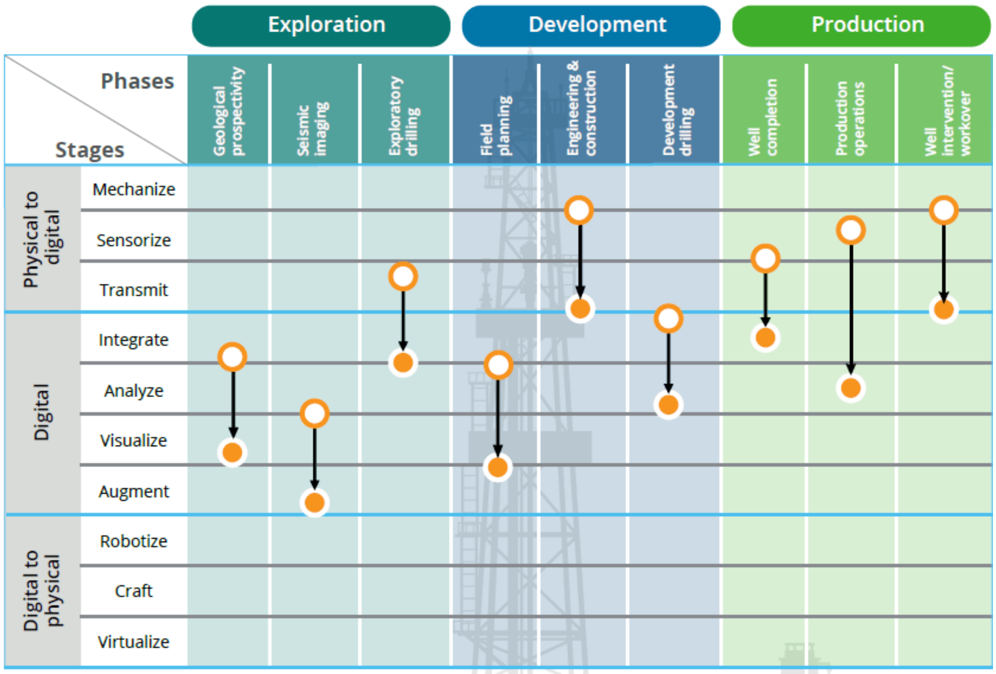

Figure 4: Digital maturity for Upstream operations. The dots represent the state as of 2017 and 2020. Exploration is ahead with analysis and visualization, whereas production and remedial operations are undergoing the switch from analog workflows to digital integration [13].

4. Petroleum Engineering

4.1. Oil and Gas Industry

The oil and gas industry is split up into the upstream, midstream and downstream side of the business. Those three sectors combined form the complete value chain from finding hydrocarbons, production, transportation, processing and ultimately the sale to customers as for industrial uses and fuel products. A vertically integrated oil and gas company (IOC) operates in all three sectors from "the well to the tank".

Downstream is the petrochemical segment with the refining and distribution of oil, gas, and fuel to customers. It includes the commercial side of the business to market and distrubutes to the end-users. The final products can be fuels for transportation, power generation, and utilities, industrial gases, and chemical feedstock.

Midstream operates the energy infrastructure to gather, store and transport hydrocarbons. This can be local or global with trucks, pipelines, tankers and LNG (liquified natural gas) carriers to connect the hydrocarbon producers to its international markets.

The main area for petroleum engineers it the **Upstream sector**, also called Exploration and Production (E&P). It includes all the steps from finding hydrocarbons, drilling wells, and its exploitation. The Upstream sector can be further split up into phases where the different disciplines of Petroleum Engineering are positioned. PE education trains engineers mainly for the Upstream sector and usually consists of specializations for reservoir, production, drilling, and economics.

The fundamentals and skills for all areas are usually taught in undergraduate courses together with a solid understanding of STEM basics. In the industry, there is the same distinction for different job roles and departments. A common problem is efficient communications across individual disciplines among the big and hierarchically organized oil and gas companies. This compartmentalization led to "silos" where data is stored in different quantities and qualities with non-uniform interfaces and standards.

4.2. Reservoir Engineering

A reservoir engineer prepares estimates and predictions of the subsurface hydrocarbon reservoirs. Based on the reservoir fluids (oil, gas, and water) and rock properties it can be determined how much of the reserves are recoverable and what the optimum methods to do so will be (field development planning). The aim is to maximize the amount of produced oil and gas and do so quickly to recover investment costs and make a profit. After the first appraisal wells are drilled there will be a confirmation if oil and gas are present and how extensive the reservoir is.

For conventional resources, numerical simulations are the tool of the trade to maximize the ultimate recovery by modeling the fluid flow and interactions underground. Unconventional resources, like shale and heavy oil, are atypical so the existence and extent are usually known beforehand, but finding an optimum completion strategy is more involved as the hydrocarbons will not flow without stimulation activities.

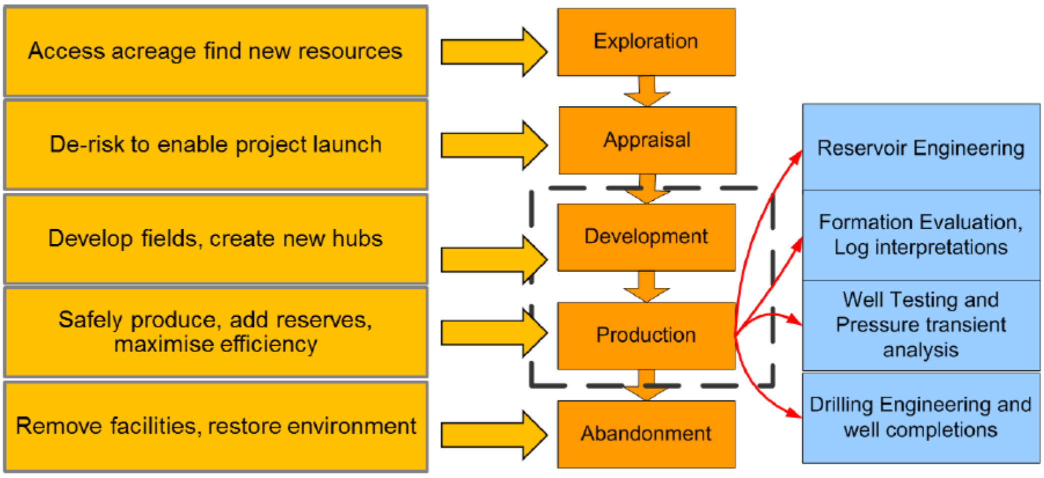

Figure 5: Lifecycle of the Upstream oil and gas industry (E&P) [7].

4.3. Drilling Engineering

Drilling Engineering is the hands-on side of PE, where a physical connection is made between the surface and the reservoir. Basically, a hole is drilled and steel pipes mounted in a well to be able to produce oil and gas. Since a hole is made into a reservoir with combustible and pressurized fluids the drilling engineer has to make sure the operations are within the mechanical limit and the pressure envelope of the equipment and geologic strata it passes through. The main goal is to drill a quality well in a safe and fast manner, as time ultimately determines the overall costs due to the high rental rates of specialized equipment and services. A disruptive technology, which allows for widespread shale developments, is horizontal drilling that multiplies the contact area within the reservoir (pay zone). For conventional oil and gas developments drilling a well is the biggest cost driver, especially offshore in deep water or remote locations. On the other hand, for unconventionals, the completion phase became the leading expenditure as it covers consumables, equipment, and services for multiple hydraulic stimulation stages commonly performed within a single well.

4.4. Production Engineering

Production Engineering is the next stage after drilling a well and the engineer specifies the equipment needed for production along the long lifecycle of an oil field. This includes the inner tubing installed in a well to allow the flow of reservoir fluids to the surface. Important parameters are its diameter, as it determines the ability to flow, and material selection to counteract corrosion and minimize repairs and remedial actions. Installation and selection of pumps, safety equipment and disposal of formation water are within this domain. Production engineers make sure a well can flow and keep the operations running with minor interruptions for routine maintenance, repairs, and upgrades. Eventually once a well has reached its lifetime it will be abandoned. Equipment, such as pumps, pipes, and valves, are removed and the site made safe and permanently isolated from the reservoir. The onsite work is usually executed within the drilling engineering domain by well servicing units or external service providers.

5. Data Analaytics

Data analysis has been around for many years and its underlying concepts and processes are well established and taught in STEM disciplines, building upon fundamentals in mathematics and statistics.

> *"Data analysis is a process of inspecting, cleansing, transforming and modeling data to discover useful information, informing conclusion and supporting decision-making. Data analysis has multiple facets and approaches, [...] In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively."* [14]

What has changed today is its widespread use among engineering disciplines, following its broad adaptation in finance, business, medical and scientific research. This is partly caused by the progress made in information technology and capable computing hardware that make acquisition and processing of vast amounts of data possible for virtually everyone. What it means though is that engineers will need a certain skill set and broad understanding of the underlying principles to understand the language of data science and make a decision upon statistical results or probabilistic predictions.

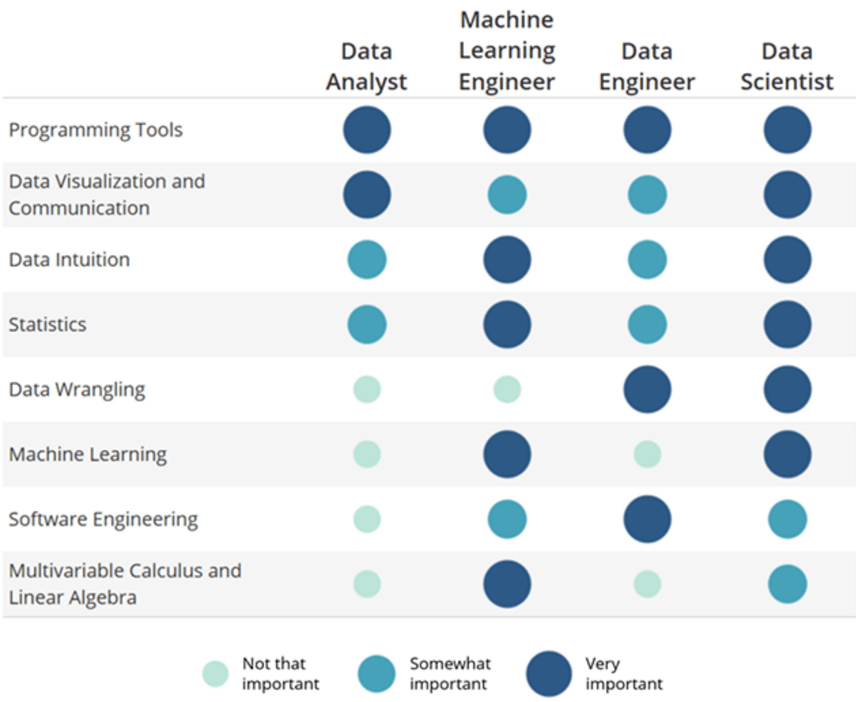

Figure 6: Distinction of skill levels for different roles within data science. Engineers, such as PE, fit the role of digital analyst and digital engineer, with emphasis on software engineering and programming [15].

UC Berkley defines the following roles in its data science degree [16]:

- **Data scientists** examine which questions need answering and where to find the related data. They have business acumen and analytical skills as well as the ability to mine, clean, and present data. Businesses use data scientists to source, manage, and analyze large amounts of unstructured data. Results are then synthesized and communicated to key stakeholders to drive strategic decision-making in the organization.

- **Data analysts** bridge the gap between data scientists and business analysts. They are provided with the questions that need answering from an organization and then organize and analyze data to find results that align with high-level business strategy. Data analysts are responsible for translating technical analysis to qualitative action items and effectively communicating their findings to diverse stakeholders.

- **Data engineers** manage exponential amounts of rapidly changing data. They focus on the development, deployment, management, and optimization of data pipelines and infrastructure to transform and transfer data to data scientists for querying.

The role of a petroleum engineer is to interpret results from geology and petrophysics and derive a development plan for commercial oil and gas extraction. The role of a data analyst is similar in that it translates technical input and compiled data into a business context. Likewise, digital engineers should be able to analyze data as part of their daily job for the specific disciplines they are experts in. As such the data handling and analytics skills needed are similar to data science, although the level can vary from awareness, knowledge or mastery for specific tasks.

The oil and gas industry with their long business cycles lead to slower adoption of emerging technologies than other industries. Still, exploration and production (E&P) were among the pioneers with modeling, big data processing, digitalization, and remote operations. A mixture of state of the art and several generations of IT/OT (information and operational technolgies) and process control equipment is in use at the same time with different implementations among assets and countries. The latest industry downturn accelerated the adoption of data science methods to optimize and analyze capital intensive Upstream projects and day-to-day operations. A sense of urgency increased the pressure to act and become on par with other industries in terms of revising established procedures and streamlining internal processes. There is still a lot to gain as these changes happened quite recently, but it is worth noting that E&P operations became profitable at a much lower oil price within a business year.

The second driver for the adoption is the boom of fracking operations, especially in North America. An enormous number of new wells are drilled and substantial investments made by many small to midsize operators in addition to the big international oil companies (IOC). Business Intelligence (BI) and data analytics are used to evaluate the economic value and make production forecasts of assets under development. New completion strategies are tried out and implemented quickly to maximize recoverable reserves and minimize costs with a steep learning curve. Reservoir engineers make use of data science to optimize field development plans and well parameters, while the underlying physical concepts and mechanisms are not yet fully understood.

Last but not least the next generation of engineers grow up as digital citizens and an advanced understanding of data science approaches is needed to interpret results and incorporate them into engineering and build interfaces to the physical infrastructure in E&P.

The basic steps in data analysis, following the input-process-output model (IPO), are:

1. Data acquisition and collection.
2. Data preprocessing and cleaning.
3. Data analysis and presentation.

The *Cross-industry standard process for data mining* (CRISP-DM) standard describes six different phases in its open standard for data mining. Figure 7 illustrates the typical steps of a data analytics project with its methods and techniques.

- Before starting an analysis the goals and business **objectives are defined** to guide the analysis and frame the problem to be solved.

- Next is the **collection** of available data and identifying its sources. Data can be internal, within an organization, or external such as open databases or commercial data providers, especially for specialized industrial information and business context. Numerous representations for structured (SQL, CSV), semi-structured (XML, JSON) and unstructured data exist (PDF, TIFF, images and documents).

- For **processing** the data is combined in a data lake, an environment to include the different formats in a NoSQL database. An important and time-consuming step is to **prepare** the data for analysis by transforming the individual sources into a standardized format.

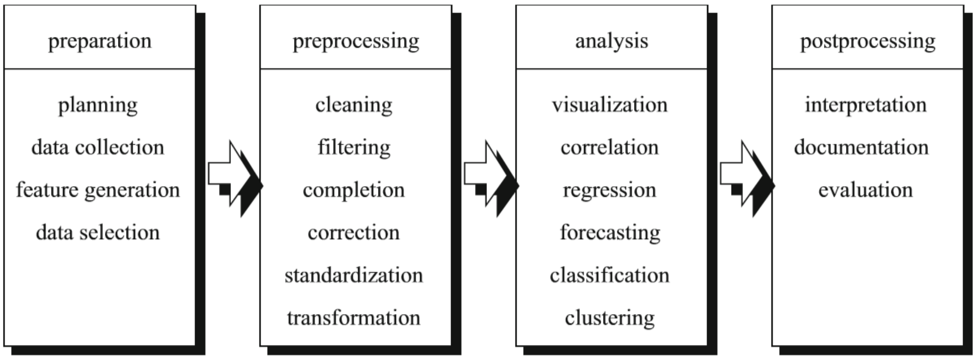

Figure 7: Typical phases of performing data analytics. The analysis and interpretation can be split up further, but all representations consist of at least an input, processing (analysis) and output phase [17].

- A central phase is **cleaning** of the data, as mistakes will generate incorrect results. Gaps, spelling, coding, duplicate entries, and errors have to be identified. This process is done semi-automatically with statistical methods for larger amounts of data with validation (sanity checks) by the analyst.
- The **analysis** itself employs different methods from math, statistics, and computer science to perform the particular analysis, for example qualitative, quantitative or exploratory data analysis (EDA).
- Finally, the results can be **interpreted** and forecasts or recommendations made. Usually, the key results are visualized to **communicate** the findings. Care has to be taken not to over simplify or convey wrong or incomplete findings. Also, the results and process should be integrated into existing workflows and reporting environments (data management) [17].

The following skills were identified for the role of data analyst [16]:

- Programming skills (SAS, R, Python).
- Statistical and mathematical skills.
- Data wrangling and data visualization.

5.1. Data Science

For engineering data science offers benefits beyond pattern recognition and in-depth analysis, as it allows to model and describes complex systems. In addition to experimental and physics-based models, it is possible to derive data-driven models to speed up computational expensive tasks or describe processes not fully understood or too convoluted to simulate. Hybrid approaches, combing physical models with data science, are widely used in geophysics, reservoir engineering, reliability, and forecasting [18].

6. Digital Transformation

At the 2018 PE Education Colloquium Prof. Donald Paul summarized the digital transformation for E&P in four phases [20]:

1. Surveillance, monitoring, incident and event response.
2. Data analytics, preventive actions, predictive operations, and cyber-security.
3. Automation of the oil field "factory", building in IIoT with open process control systems and standards.
4. Autonomous operations .

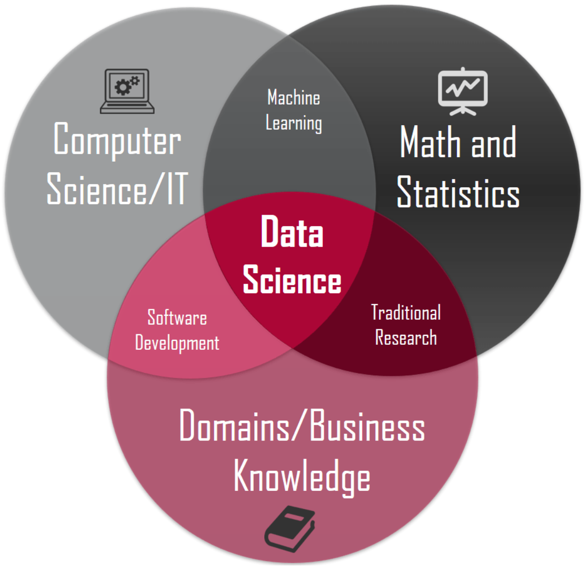

Figure 8: Venn diagram showing the multidisciplinary nature of data science. For engineering, such as PE, the domain knowledge is essential due to its specialized nature and the red circle is analog to digital engineers with applied IT and math skills [19].

Phase 1 is the Digital Oilfield and a reality for Upstream operations worldwide. Currently, E&P is implementing phase 2 and setting standards for the Digital Oilfield 2.0 in phase 3. Refining and the petrochemical industry is well ahead of E&P in terms of digitalization and optimizations, as efficient operations are key in Downstream with lower profit margins. Autonomous operations, the final phase 4, is the goal and vision for Industry 4.0 with extensive R&D efforts for every aspect of manufacturing and services [20].

6.1. Industry 4.0

Industry 4.0 is a term used to describe the effects of digitalization on the manufacturing industry and how high-tech changes the industrial landscape. The term was introduced by the German Government for its digitalization strategy in 2013. Similarly themes and digitalization strategies for the "Fourth Industrial Revolution" are legislated by all economies, such as *Made in China 2025* (China), *A National Strategic Plan for Advanced Manufacturing* (US), *The Future of Manufacturing* (UK) or *White Paper on Manufacturing Industry* (Japan) [21].

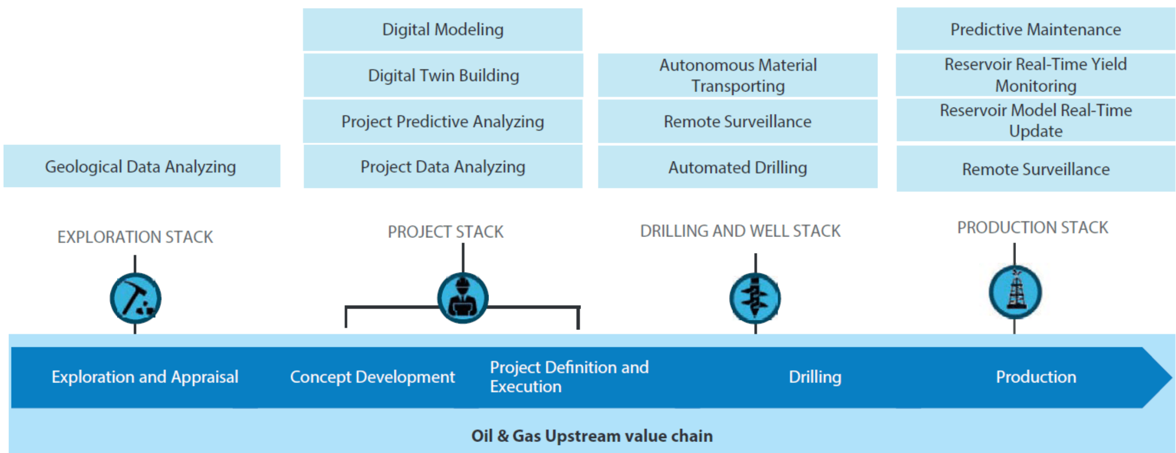

Figure 9: Applications of Industry 4.0 solutions within Upstream E&P [22].

6.2. The Digital Oilfield

The Internation Energy Agency (IEA), formed after the 1973 oil crisis as think tank for oil and energy sector, describes digitalization as follows [23]:

> *"The oil and gas sector has a relatively long history with digital technologies, notably in upstream, and significant potential remains for digitalization to enhance operations further. The widespread use of existing digital technologies could decrease production costs between 10% and 20%. With the use of both existing and emerging digital technologies, technically recoverable oil and gas resources could be boosted by around 5% globally. The potential impact of digitalization is likely to be greatest for tight oil and shale gas resources."*

Figure 4 shows the overall maturity and adoption of the digital transformation within the oil and gas industry. The adoption of new technologies, streamlining processes and optimizing its operations became a necessity as a result of industry downturn in wake of sinking oil prices and "lower-for-longer" strategies to become profitable with positive free cash flow.

Digitalization is not new in the industry and became known as Digital Oilfield (DOF) in the late 1990s. A broad adaption of high-performance computing in E&P happened at the turn of the millennia for seismic processing, visualization and numerical simulations of reservoir fluid flow. As such geophysics and seismic are among the pioneers of working with massive data sets and supercomputing for industrial applications.

After exploration, the next wave of digitalization was with drilling and downhole monitoring. Measurement-while-drilling (MWD) and logging-while-drilling (LWD) is the real-time transmission of sensor data and measurements while a well is being drilled. This was one of the key enablers for horizontal drilling, as the location and orientation could be continuously monitored and the drill bit reoriented to adjust the well path. Digital telemetry makes it possible to have a better understanding of what happens downhole and to remotely monitor data from sensors in the field during production.

In 2015 the third big wave happened following a downturn and the need to do more with less, including reevaluation of the wealth of already existing data and implementing Industry 4.0 techniques to increase efficiency and internal communications [24].

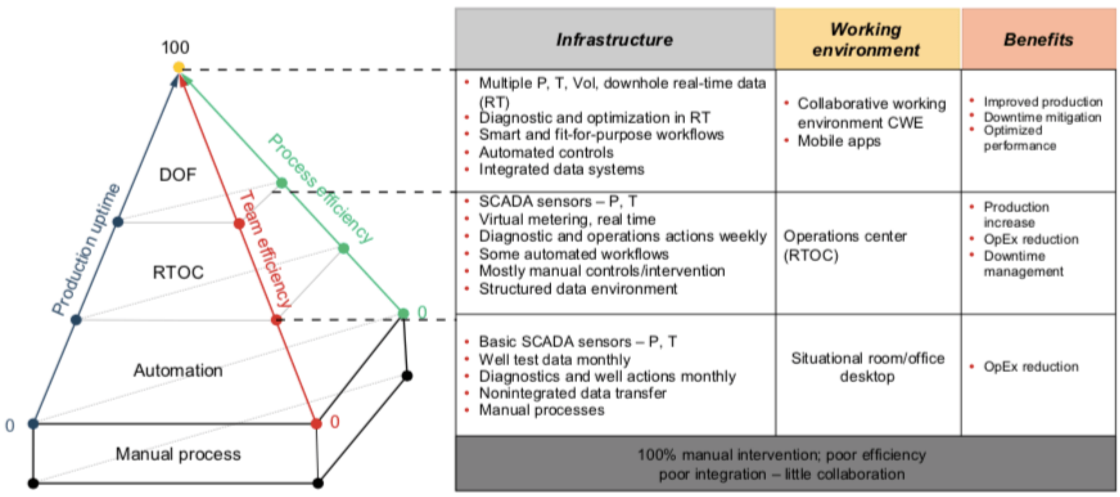

Figure 10: Four operational levels of the Digital Oilfield (DOF). The column *Infrastructure* list examples of the integration and deployed technologies [1].

Data is one of the key enablers for DOF. The complexity of production systems increased as did the granularity and amount of data collected every day. Information management remains a major challenge as operational personnel and engineers are absorbed in their daily activities in addition to collecting and filtering suitable data. Compartmentalization among different disciplines and business units discourages information flow between processes that should be more integrated. Systems and facilities evolved and different silos continue to exist when assets are acquired from other companies or technologies from various vendors are employed in the field.

The components of DOF technology are [25]:

- IT Infrastructure
- Field Instrumentation and Control
- Workflow Automation
- Multi-disciplinary Collaboration

6.3. *Digital Oilfield 2.0*

Currently, *digital transformation* is a buzz word among almost all industries as is the adaption of digital solutions and data science methods. Within E&P it is called Digital Oilfield 2.0 and changes the way of working with the wealth of available data. Another challenge is the knowledge transfer to the new generation of engineers, as the "big crew change" is happening and a big population of experienced workers retires. Blended training, online courses, and lifelong learning will become enablers to keep up with technologies and effective transfer of knowledge [26].

6.4. *Industrial Internet of Things*

The Internet of Things (IoT) is the connection of everyday devices and sensors over the internet. For industrial uses, IIoT can be seen as an expansion to SCADA (supervisory control and data acquisition) allowing access over public networks without special hardware or physical vicinity. Modern wireless communications allow connecting virtually any devices to the internet, data transfer and permanent storage of large amounts of data. The first DOF applications were connecting existing systems to external networks and implementing real-time data access and information dashboards outside of on-site control rooms. Real-time operations center (RTOC) were the forerunners of IIoT in the industry before wireless technology and widespread internet access became a commodity. Paired with centralized storage and computing infrastructure it is possible to store and access high-bandwidth data to use for analytics involving complex processing. Within industries, such as oil and gas, a plethora of data is already collected but was not made available outside of a limited scope or simply discarded [25].

Integrated Asset Management	**Real-time Production Optimization**	**Real-time Drilling and Completions**
• Advanced geology and integrated interpretation platforms • High performance computing center for multiscenario analysis and multi-million cells reservoir simulation • Automated reservoir history match • Real-time down-hole pressure and rate analysis • Real-time reservoir surveillance practices using seismic-based, electro-magnetics, tracers, streamline simulation and data-driven • Smart waterflooding • Integrated Forecasting • Integrated planning and scheduling • Automated asset portfolio management • Surrogate and proxy modeling of subsurface	• Automated workflows for well test validation, well model update, and plant update. • Real-time KPI calculations • Continuous plant efficiency, uptime and availability • Real-time integrated production optimization • Virtual metering: multiphase flow rates, soft sensors, and zonal allocation • Smart well monitoring and control • Smart diagnosis and by exception-base surveillance, expert alarming • Predictive advisory short-term forecast • Artificial lift diagnosis and optimization • Immersive collaboration environments • Remote operating centers • Operations dynamic simulator • Advanced process control	• Remote operating centers • Remote geo-steering • Anti-collision optimization • Real-time weight on bit optimization • Real-time fracture optimization • RFID-based for downhole completions control • Smart diagnosis and by exception • Operations dynamic simulator

Figure 11: Discplines and application of DOF and IIoT [25].

Cybersecurity risks scale together with the benefits when connecting physical systems over a public network and enabling access to internal processes and business data over the internet.

Following statement sums up the advantages of the Industrial Internet of Things (IIoT):

"The IIoT can greatly improve connectivity, efficiency, scalability, time savings, and cost savings for industrial organizations. Companies are already benefitting from the IIoT through cost savings due to predictive maintenance, improved safety, and other operational efficiencies. IIoT networks of intelligent devices allow industrial organizations to break open data silos and connect all of their people, data, and processes from the factory floor to the executive offices. Business leaders can use IIoT data to get a full and accurate view of how their enterprise is doing, which will help them make better decisions." [27]

6.5. *Cyber-Physical Systems*

Cyber-physical systems (CPS) allow integrating physical processes with complex computation. CPS goes beyond robotics and embedded systems as it allows to dynamically interact with the physical world. As such CPS reacts to the physical world and vice versa to build an autonomous system, such as self-driving cars or unsupervised drones performing inspections and maintenance. As part of Industry 4.0 expectations are to alter manufacturing by automating processes and enabling to scale them up rapidly (smart factories). Within oil and gas systems for automated drilling, maintenance in remote areas and inaccessible locations are being developed and prototyped in the field.

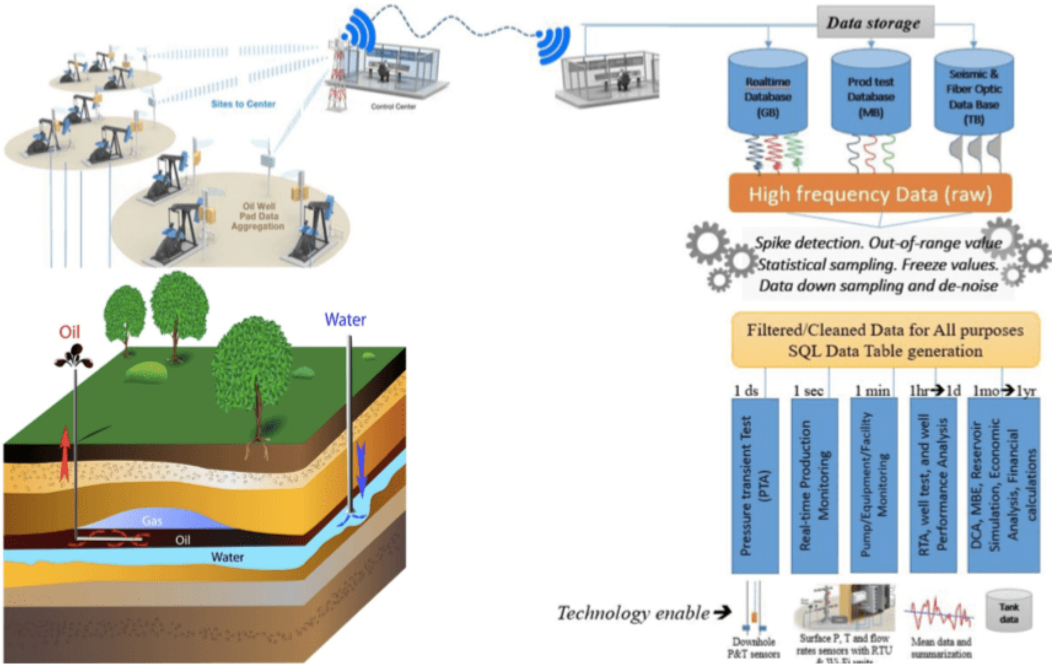

Figure 12: Real-time operations within the Digital Oilfield. The right side depcits centralized processing (RTOC of Cloud solutions) and typical analytics performed by Petroleum Engineers [1].

6.6. *Digital Twin*

A digital twin is the virtual clone of a physical asset, machine or plant. Usually, it is represented as a 3D model of the actual site as a human-machine interface (HMI) to access and display real-time sensor data and operational parameters. Moreover working with digital twins allows to optimize and simulate processes, on a plant or asset level, before machinery is constructed or put in operation. The benefits lie in the availability of advanced analytical and monitoring capabilities to predict and simulate the behavior of whole systems. Data science can be used to analyze complex correlations and identify patterns for predictive maintenance, bottlenecks, and operational risks. A reservoir simulation can be seen as digital twin of the physical reality and is used to predict an optimum development strategy. Upstream operations are very expensive and there are limited options to get it right, as wrong choices early can impede production years later and reduce the total amount of recoverable reserves.

6.7. Cloud and Edge Computing

Cloud computing is a server-client infrastructure where resources, such as processing power and data storage, are allocated on-demand. Instead of operating dedicated data centers and server farms hardware is rented from specialized cloud providers. Resources are used and scaled up as required, maximizing its utilization and saving on operating and investment costs. A big advantage of cloud computing is the sophisticated analysis and visualization tools made available from the cloud providers software as a service (SaaS). The leading companies invest heavily in data science research and make their tools and technologies available to the customers, allowing easy access to state-of-the-art and user-friendly data analytics tools.

Edge computing is different to cloud computing as processing and storage are performed on premise where the data is acquired. This allows bandwidth savings to transfer large amounts of data to a centralized cloud, such as raw data from IoT enabled sensors or video feeds. The latency and reaction time is improved as computation happens closer to the actual activities. The processing power and volume of data from embedded systems increase exponentially. Advanced techniques, like speech recognition or image recognition, are already deployed on the edge in consumer devices like mobile phones.

6.8. Big Data

Big Data refers to the acquisition, storage, and processing of massive amounts of data. It includes structured and unstructured data in a variety of different formats. Big Data can be described by several characteristics such as its volume, variety, quality, speed, and value. Business intelligence is a subset of Big Data which analyzes, predicts and visualizes trends, whereas data science allows disovering new relationships [28]. In general Big Data requires infrastructure for cloud computing or data centers, as its volume is too big to store on workstations and necessitates parallel processing on multiple servers.

Within oil and gas, huge amounts of data are collected, however, only a small part is actually used to be analyzed and the majority discarded or ignored due to practical reasons. Concepts of Big Data, first applied in retail and social media industries, are already changing engineering and will have a big impact on E&P due to the sheer amount of data collected and becoming accessible in a structured format. Data analysis is core to the extraction of information and interpretation of probabilistic results. The fundamental difference to traditional approaches is collecting large volumes of less specified information and compensate disorganized ("messy") data with its sheer volume [29].

Feblowitz summarizes Big Data in Upstream Oil and Gas as follows [30]:

- *More sophisticated technologies are producing more seismic, drilling and production data than ever before.*
- *There are some credible potential applications of Big Data and analytics to deliver benefits in exploration, development, drilling, production operations, maintenance, and the enterprise.*
- *Many of the challenges of accessing data have been overcome with faster networking technologies, but there are still input/output storage challenges such as insufficient capacity, limited upload bandwidth, and multiple access patterns.*
- *Work on the application of Big Data and analytics in the oil and gas industry is in the experimental stage, with much of the early work focused on data-intensive computing and how I/O data loading can be managed most efficiently.*
- *Oil and gas companies that want to take advantage of data to gain a competitive edge and lower risks should formulate a Big Data strategy that includes the evaluation of decision-makers' requirements, decision processes, existing and new technology, and the availability and quality of data.*

7. Method

7.1. Engineering Learning

A core goal is to provide students and teaching faculty with a coherent, well-designed and engaging learning experience [31]. The Colorado School of Mines uses its own *Engineering Learning at Mines* framework for designing and endorsing coursework. The complete process is illustrated in figure 13:

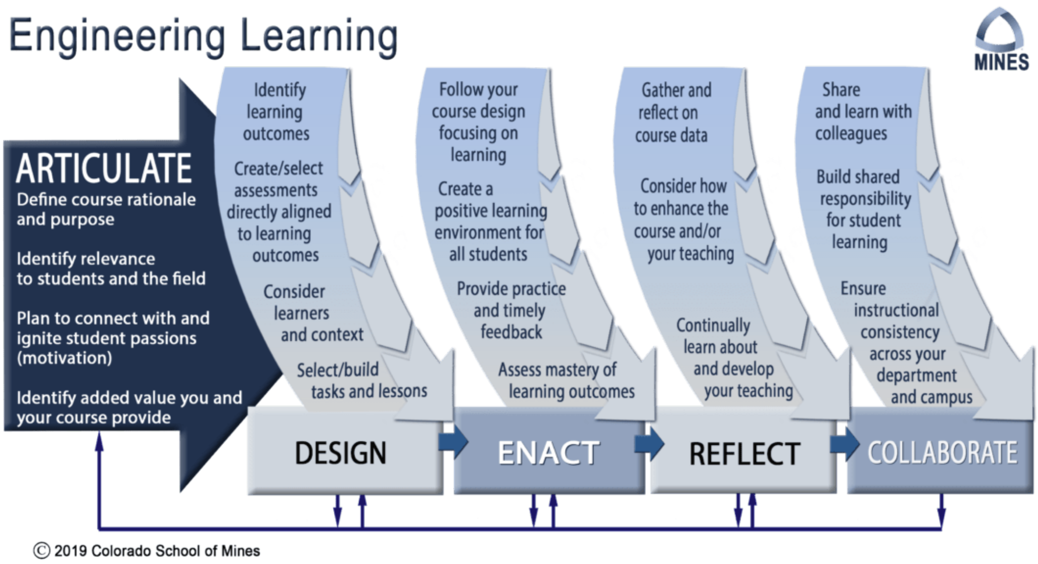

Figure 13: Course design process at Colorado School of Mines, Golden CO [31].

- **Articulate**
  The first phase is to formulate the purpose of the course and the rationale, including its relevance to the field of study and where it fits into a degree program and what the prerequisite classes are. The added values take a holistic view of applied skills and knowledge that will be conveyed to students.

- **Design**
  The next step is a top-down approach to verbalize overall learning outcomes (LO) for the course. This can be formulated as SMART (specific, measurable, achievable, realistic, and timely) goals, with emphasis on being specific and measurable. Bloom's taxonomy or Webb's Depth of Knowledge is used in defining the level of understanding expected after completing the course [32]. Mastery and critical thinking are anticipated at the graduate level: the highest stage on those scales.

- **Define**
  Defining the outcomes leads to specific tasks and forms of assessments, which will be itemized for the individual lessons or modules to gradually increase the knowledge to the desired LOs for successfully finishing the course.

- **Enact**
  The individual lessons are designed during the enact phase. The LOs are specified and defined for separate learning activities. To positively engage students different methods of teaching can be used and interactions with and among students promoted. Care has to be taken to address every student and individually support their talents. During assessments, students will demonstrate their mastery and refine their knowledge when solving real-world problems.

- **Reflect**
  Reflections are done by an instructor after teaching a course and noting what went well and what did not. A set of questions helps to formulate feedback and prepare refinements for the next module or lecture.

- **Collaborate**
  Collaboration is a review with colleagues, as the course is developed (the LOs defined) and what approaches work for teaching. It gives also an opportunity to align courses collectively and tailor them for each college.

7.2. *Online Learning*

One main goal was to make the course available to a wider audience, including industry professionals and students off-campus. For this reason, the courses were developed as online courses in *Canvas*, the learning management system (LMS) used at Mines. The data analytics coursework utilizes different online features, such as discussion boards, video lectures, live video, screen sharing, online collaboration, library access and secure VPN access to hosted software and databases.

It is noteworthy that some of the best data science courses available are pure online courses and available for free or a small fee as Massive Open Online Courses (MOOC). For data science and PE distance learning and online courses already exist from different providers, but no combination specialized for incorporating modern data approaches specific for Petroleum Engineering.

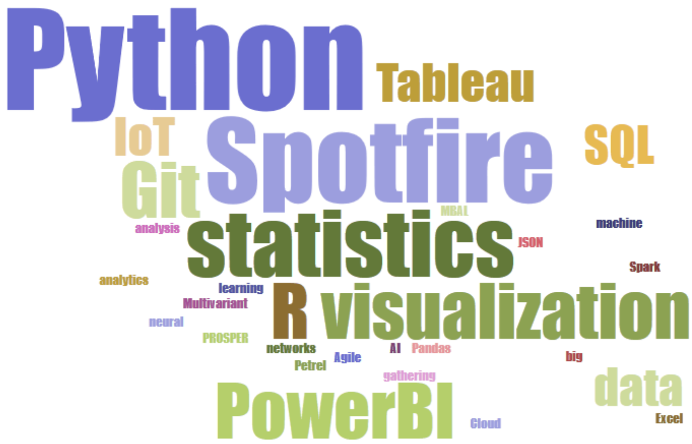

Figure 14: Word cloud from our survey about topics and specific technologies to prepare students for the digital engineering environment within PE.

A survey was conducted among E&P professionals and Mines graduates about the expectations, skillset and recommended tools to be relevant in the industry for the current and upcoming graduates. See figure 14 for a graphical representation of the free-response questions regarding central topics and technologies.

The proposed course outline focus on domain-specific knowledge and the current toolset used within the oil and gas industry and builds upon graduate level fundamentals in math and computer science.

7.3. *Mines Online*

Online coursework to be offered at Mines has to go through accreditation from the regional *Higher Learning Commission* in addition to the regular review and approval process [34]. One more difference is that the whole coursework, including the lectures and assignments, needs to be completed ahead of time for revisions and quality assurance. See figure 13 for the design and development process for online classes [33]. During its first rollout, close attention will be paid to establish an engaging and interactive learning experience for the students, as there will be no face-to-face sessions for a fully online course.

8. Results

8.1. *Graduate Certificate*

For Petroleum Data Analytics a new online curriculum is developed[1]. It is based on the exiting undergraduate PE Data Analytics Minor taught on campus [35]. More advanced projects and applications were added for graduate-level course work. The proposed curriculum consists of four courses with three credit hours each:

[1]The results presented here are a prelimnary outline and subject to change for the final syllabus.

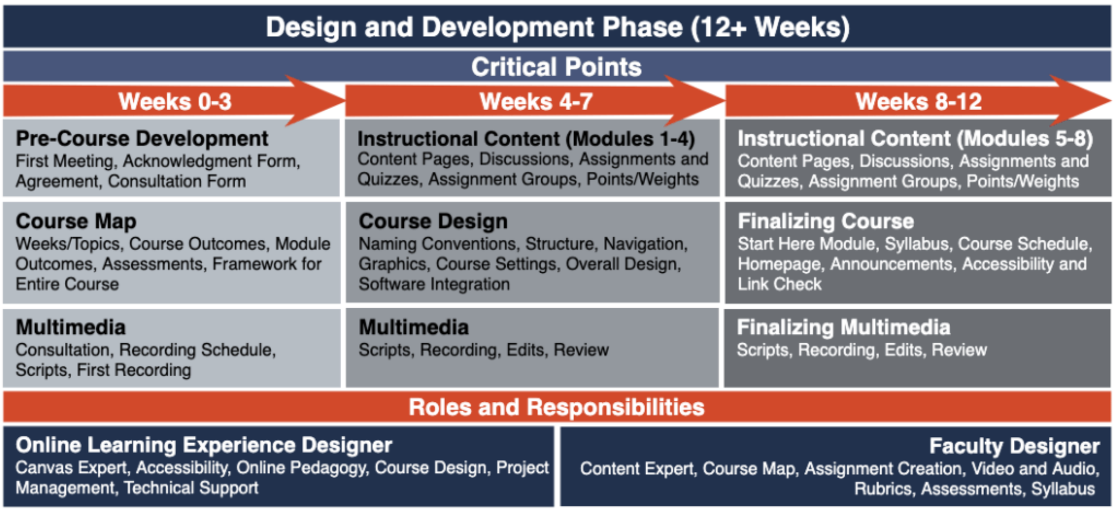

Figure 15: Phases for online course design. Preparation and coursework development take at least three months in addition to the approval and accreditation process [33].

- **Introduction to Data Science**

  This course is taught with the help of Python and Jupyter notebooks. Programming in Python is introduced and applied to learn and apply visualization, fitting, data acquisition, classification, clustering and web scraping. The concepts of data science are explained and an introduction to neural nets and machine learning is given.

- **Advanced Math and Statistics**

  Introduction to probability, random variables, and discrete and continuous probability models. Elementary simulation and bootstrapping. Data summarization and analysis. Confidence intervals and hypothesis testing for means and variances. Distribution-free (i.e., nonparametric) techniques.

  Statistical computing can be done with free software (R, Octave, Python) or commercial packages (Matlab, Spotfire).

- **Petroleum Data Analytics Fundamentals**

  This course builds upon the data science and math classes with Petroleum Engineering domain knowledge expanding on the specifics covered in this report.

  1. Introduction to the Digital Oilfield 2.0, what is new today, career opportunities.
  2. Operational technology (OT) and information technology (IT) systems, field network, SCADA, ERP.
  3. System challenges, adoption barriers, IIoT, cyber security concerns and risks.
  4. Data foundation for oil and gas, data visualization techniques, business intelligence, and Big Data.
  5. Review of statistics, analytical and visualization techniques.
  6. Information intensity in O&G, unconventional resources.
  7. Application of petroleum data analytics to Upstream data, dashboard visualization, and mapping.
  8. Digital twin, future of digital technology in O&G, integrated use cases.

- **Petroleum Data Analytics Applications**

  The applications part of PE data analytics is split up into four segments consisting of two modules each. The segmentation gives flexibility for scheduling plus the possibility to teach in parallel with the fundamentals course:

  1. Drilling and Completions

2. Production Operations and Maintenance
3. Reservoir Characterization
4. Unconventional Resource Optimization

Each segment introduces the topic with current challenges and specific problems. Data sets and examples are presented as guidance for students to compile and analyze. Each segment is concluded by presenting the workflow and findings individually or as group projects.

9. Outlook

The new courses were designed as online classes and enable people off-campus to enroll and participate. Both Petroleum Data Analytics courses consist of 8 modules each. The average workload for each module is 15-20 hours per week and can be completed asynchronously within that timeframe. This makes it possible to provide the courses to a wide audience, which fulfills the admission and enrollment criteria [36].

Additionally online courses are offered during the summer term, reducing the course load on students and faculty. Scheduling of online classes becomes easier than traditional courses as the modules are completed asynchronously with a specified timeframe. One design criteria for the curriculum was that the four courses together can be offered as a graduate certificate to professionals in the industry too.

The individual course modules are under development and the first rollout is planned for Fall 2020 or Spring 2021.

10. Illustrated Examples

Public available data and application are used for the course, the students can freely chose their prefered toolkit. Programming and statistics are taught with R and Python. Following examples are made with the business intelligence (BI) software package Spotfire that will be used for the course. Spotfire is popular within the U.S. Upstream sector and includes internal Python and R interpreters, mapping and database interfaces.

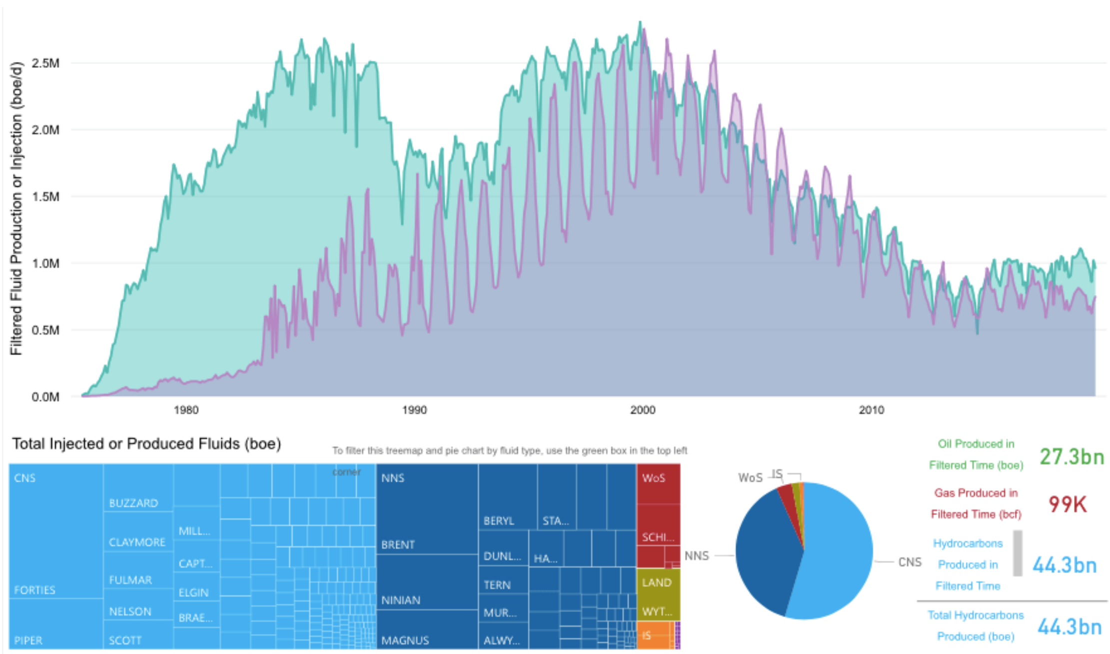

Figure 16: Current and historic oil and gas production data from the United Kingdom, realized as PowerBI dashboard from UK's Oil and Gas Authority's Open Data effort. Source https://data-ogauthority.opendata.arcgis.com/.

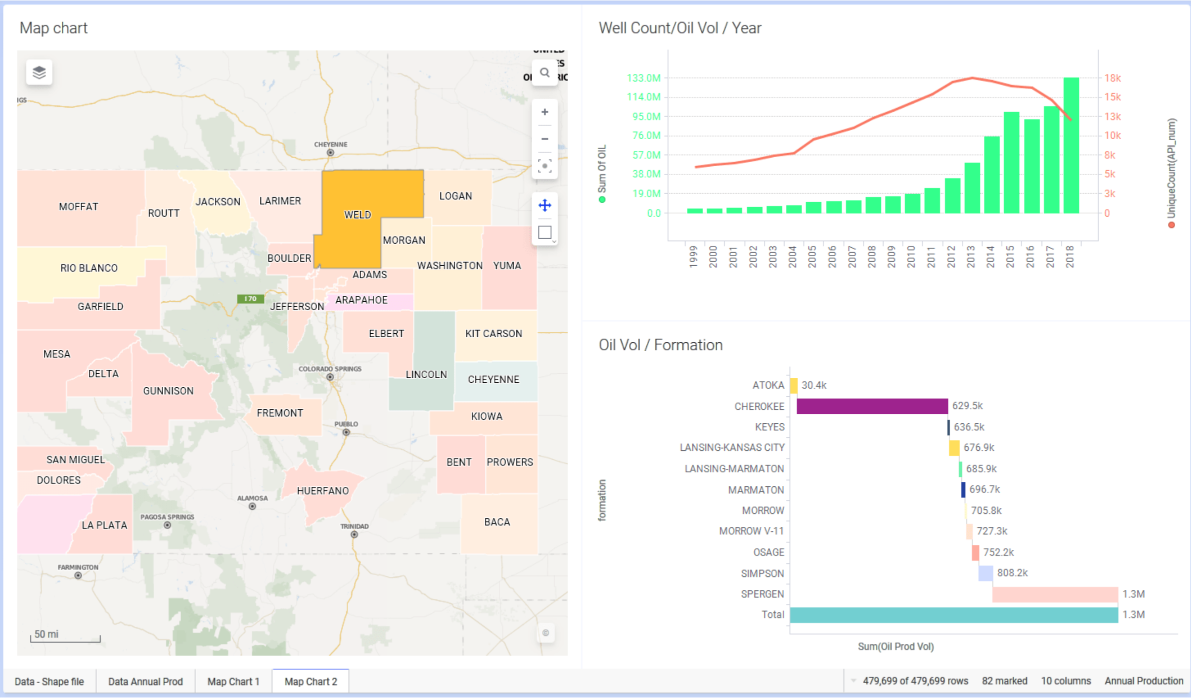

Figure 17: Dashboard for total annual production and wells drilled in Colorado per county and formation. Python is used to download and process public data from http://cogcc.state.co.us.

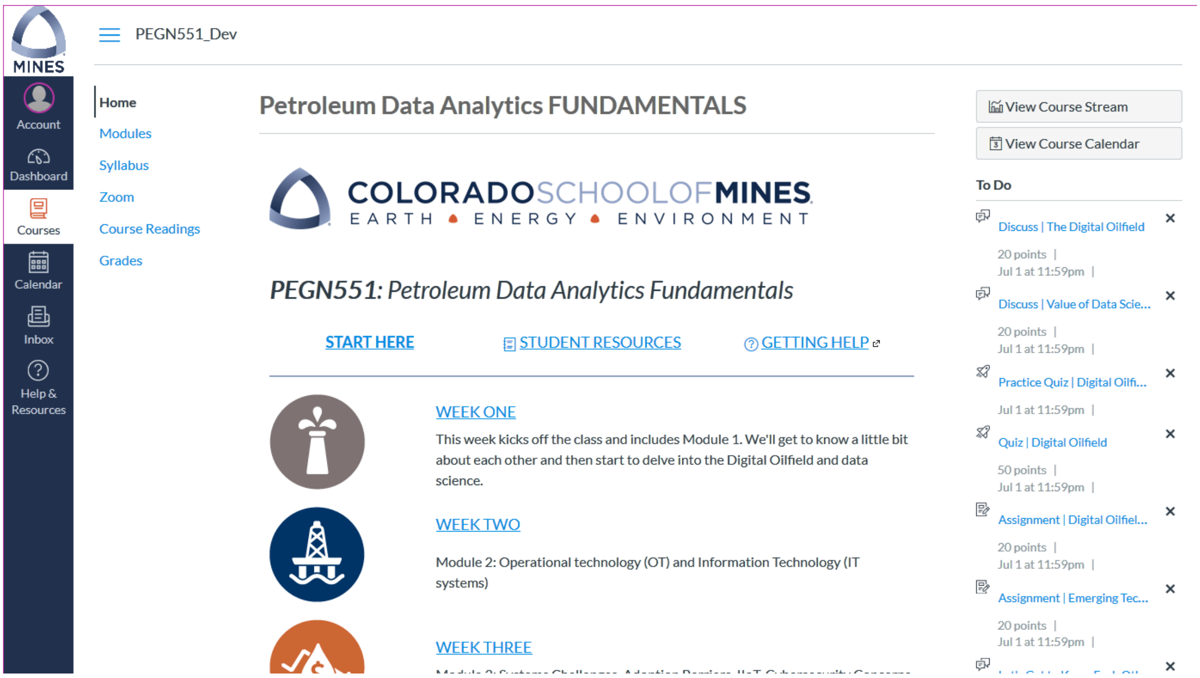

Figure 18: Student view of the interactive course in the Canvas Learning Management System (LMS).

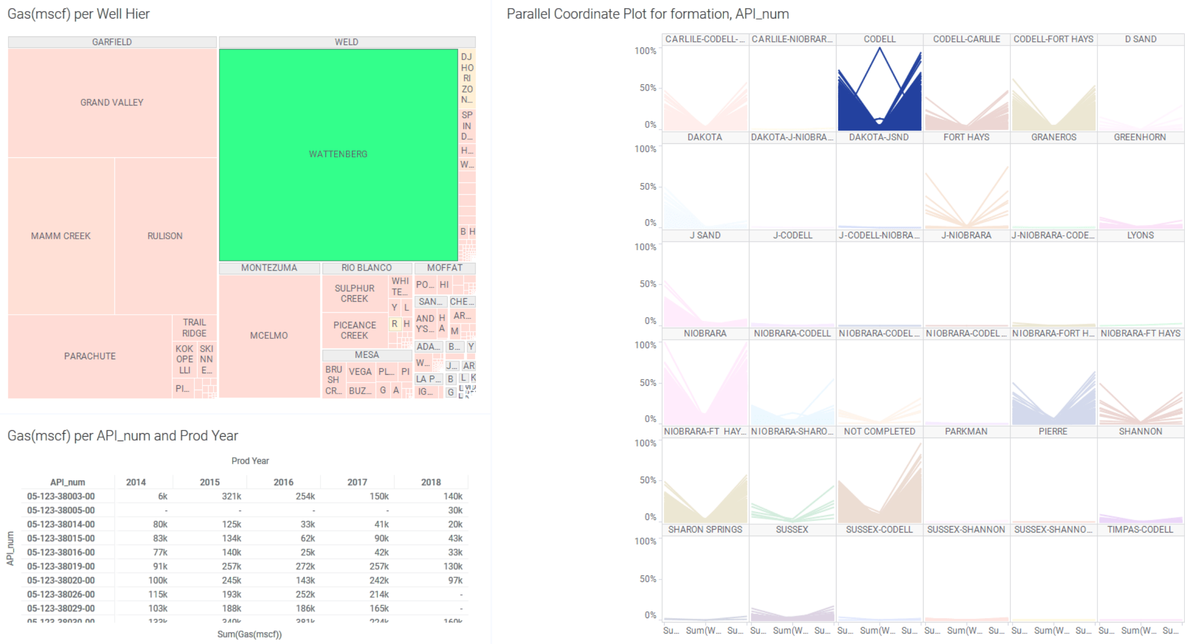

Figure 19: Normalized view of oil, water and gas production per formation (right) within a county and gasfield (top) and annual production per well (bottom), data from http://cogcc.state.co.us.

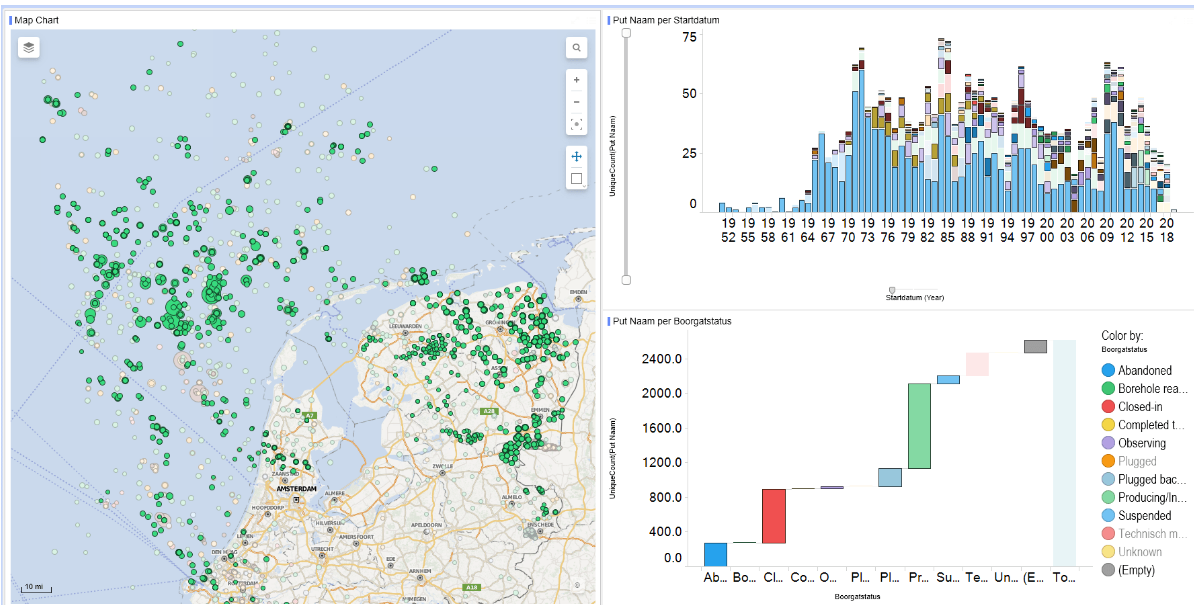

Figure 20: Map view of producing gas wells in the Dutch North Sea an onshore The Netherlands, current well status (bottom) and drilling activity per operator (top), data from https://www.nlog.nl/boringen.

Figure 21: Oil and water production data from the Norwegian Volve field, per well and year (right) and normalized to days on stream (left), data from https://data.equinor.com/volve. Decline curve analysis implemented with R code (not shown).

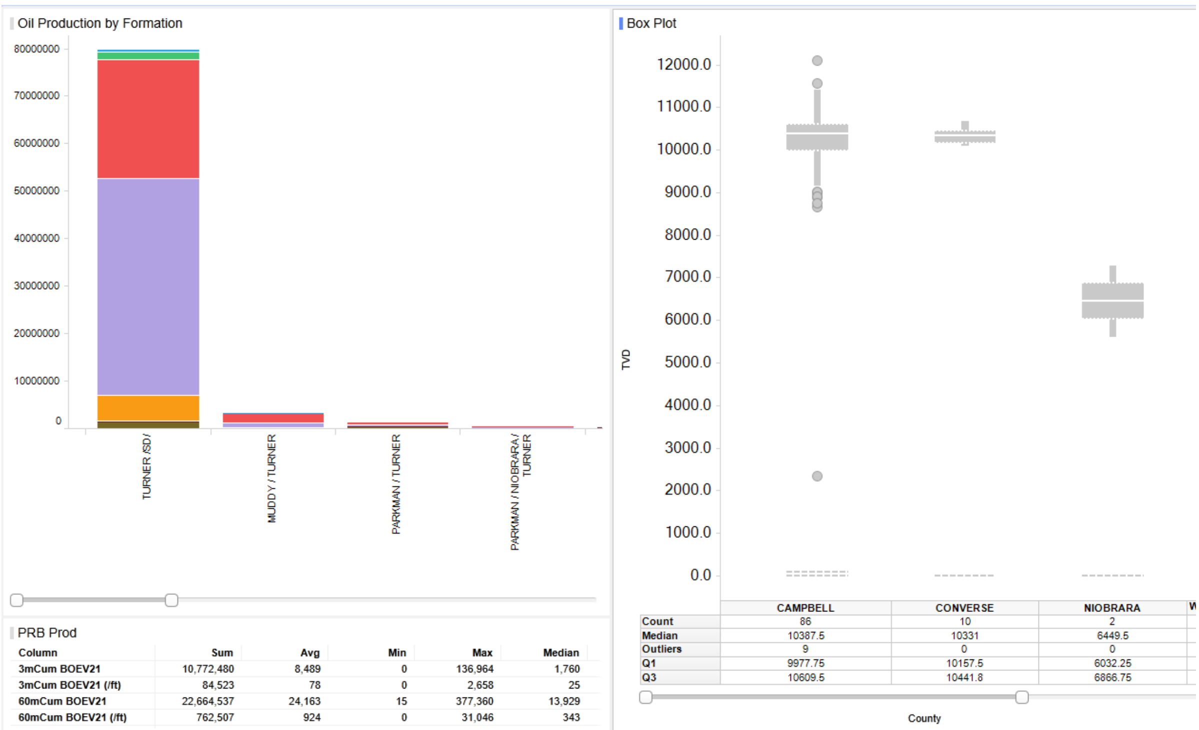

Figure 22: Analysis of the Powder River Basin in Wyoming, with statistics on perforated intervals and oil production from each formation, data from http://pipeline.wyo.gov/.

References

[1] Gustavo Carvajal, Marko Maucec, and Stan Cullick. *Intelligent digital oil and gas fields: concepts, collaboration, and right-time decisions*. Gulf Professional Publishing, 2017.

[2] Society of Petroleum Engineers. The 2018 petroleum engineering education colloquium. https://www.spegcs.org/events/3960/, . [Online; accessed: 2018-12-06].

[3] Colorado School of Mines. Internal memo, 2019-01-29.

[4] Dean Rietz, Mohamed Soliman, Adam Cagle, et al. An industry look at the petroleum engineering curricula. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2019.

[5] H Kazemi, W John Lee, Thomas A Blasingame, Rex Allman III, Zaki Bassiouni, Charles H Bowman, Alfred W Eustes, Don W Green, Lloyd R Heinze, Roland N Horne, et al. The fifth spe colloquium on petroleum engineering education-an industry perspective. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2000.

[6] Society of Petroleum Engineers. Competency matrices. https://www.spe.org/en/training/competency/, . [Online; accessed: 2019-04-10].

[7] Mohamed Hassan Sayed, John Senam Fianu, Reza Malakooti, Pablo Ariel Salazar Misslin, Michael Kenomore, and Amir Hossain Gharavi. Petroleum engineering: a collaboration between academia and industry. In *6th Annual Symposium of the United Kingdom & Ireland Engineering Education Research Network*, pages 205–218. University of Portsmouth, 2019.

[8] Nadiya Tymkiv. Curricula and programmes in petroleum engineering for higher technical education institutions: Comparative analysis. *Comparative Professional Pedagogy*, 8(1):20–26, 2018.

[9] Lloyd Heinze, Habib Menouar, Marshall Watson, Talal Gamadi, et al. Petroleum engineering enrollment: Past, present and future. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2019.

[10] Derek Mathieson, D Nathan Meehan, Jeff Potts, et al. The end of petroleum engineering as we know it. In *SPE Middle East Oil and Gas Show and Conference*. Society of Petroleum Engineers, 2019.

[11] Judy Feder. As industry changes, so does petroleum engineering education. *Journal of Petroleum Technology*, 71(12), 2019.

[12] Iraj Ershaghi, Donald L Paul, et al. The changing shape of petroleum engineering education. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2017.

[13] Anshu Mittal, Andrew Slaughter, and Vivek Bansal. From bytes to barrels. the digital transformation in upstream oil and gas. *New York, NY: deloitte insights.*, 2017.

[14] Wikipedia contributors. Data analysis — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Data_analysis, . [Online; accessed: 2019-12-16].

[15] Udacity School of Data Science. What is data science? https://blog.udacity.com/2014/11/data-science-job-skills.html. [Online; accessed 2020-01-11].

[16] Berkley School of Information. What is data science? https://datascience.berkeley.edu/about/what-is-data-science/, 2019. [Online; accessed: 2019-12-17].

[17] Mike Fleckenstein and Lorraine Fellows. Physical asset management vs. data management. In *Modern Data Strategy*, pages 15–22. Springer, 2018.

[18] Society of Petroleum Engineers. Spe workshop: Merging data-driven and physics-based models for enhanced reservoir insights and predictions. https://www.spe.org/events/en/2019/workshop/19asa2/schedule-overview.html. [Online; accessed 2020-01-13].

[19] Michael Barber. Data science concepts you need to know! https://towardsdatascience.com/introduction-to-statistics-e9d72d818745. [Online; accessed: 2020-01-13].

[20] Donald Paul. Evolving petroleum engineering education to meet the changing needs of the industry and the profession. https://webevents.spe.org/catalog?page=8. [Online; accessed: 2019-12-17].

[21] Hongfang Lu, Lijun Guo, Mohammadamin Azimi, and Kun Huang. Oil and gas 4.0 era: A systematic review and outlook. *Computers in Industry*, 111:68–90, 2019.

[22] Stuart Fraser, Thomas Anastaselos, and G.V.V. Ravikumar. The disruption in oil and gas upstream business by industry 4.0. *Infosys Limited*, 2018.

[23] International Energy Agency. Digitalization & energy. IEA, 2017.

[24] Well Data Labs. The digital oilfield – then and now. https://www.welldatalabs.com/2019/04/digital-oilfield-then-and-now/. [Online; accessed 2019-10-03].

[25] Luigi Alfonso Saputelli, Cesar Bravo, Michael Nikolaou, Carlos Lopez, Ronald Cramer, Satoshi Mochizuki, Giuseppe Moricca, et al. Best practices and lessons learned after 10 years of digital oilfield (dof) implementations. In *SPE Kuwait Oil and Gas Show and Conference*. Society of Petroleum Engineers, 2013.

[26] Joel Parshall et al. After years,'big crew change'has passed, but learning, training challenges remain. *Journal of Petroleum Technology*, 69 (07):38–40, 2017.

[27] Inductive Automation. What is iiot? https://inductiveautomation.com/resources/article/what-is-iiot. [Online; accessed 2020-01-13].

[28] Wikipedia contributors. Big data — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Big_data, . [Online; accessed 2019-12-16].

[29] Robert K Perrons and Jesse W Jensen. Data as an asset: What the oil and gas sector can learn from other industries about "big data". *Energy Policy*, 81:117–121, 2015.

[30] Jill Feblowitz. The big deal about big data in upstream oil and gas. *IDC energy insights*, pages 1–11, 2012.

[31] Trefny Innovative Instruction Center. Engineering learning. https://trefnycenter.mines.edu/engineered-learning/, . [Online; accessed 2019-10-01].

[32] Katherine Miller. Bloom's taxonomy and webb's depth of knowledge. https://www.synergiseducation.com/blooms-taxonomy-and-webbs-depth-of-knowledge/. [Online; accessed 2019-11-11].

[33] Trefny Innovative Instruction Center. Online learning. https://trefnycenter.mines.edu/whatweoffer/online/, . [Online; accessed 2019-12-16].

[34] Colorado School of Mines. Online program/course approval and development proces. https://trefnycenter.mines.edu/wp-content/uploads/sites/44/2018/06/Approval-and-Course-Development-Process_WebVersion-1.pdf, . [Online; accessed 2019-10-01].
[35] Department Petroleum Engineering. Data analytics minor. https://petroleum.mines.edu/wp-content/uploads/sites/32/2017/10/Data-Analytics-Minor-Flyer_revision2.pdf. [Online; accessed: 2019-04-10].
[36] Colorado School of Mines. Applying for graduate studies. https://online.mines.edu/admissions/, . [Online; accessed 2020-01-29].