Big Data in the Oil & Gas Upstream Industry – A Case Study on the Norwegian Continental Shelf

By G. VEGA-GORGOJO. R. FJELLHEIM, D. ROMAN, R. AKERKAR and A. WAALER*

Abstract

This case study is focused on the impact of big data in exploration and production of oil & gas in the Norwegian Continental Shelf. Overall, the industry is currently transitioning from mere data collection practices to more proactive uses of data, especially in the operations area. Positive economical impacts associated with the use of big data comprise data generation and data analytics business models, commercial partnerships around data, and the embracement of open data by the Norwegian regulator. On the negative side there are concerns regarding the future of existing business models and the reluctance of oil companies to share data. Positive social and ethical impacts include mitigation of safety and environment concerns with big data, personal privacy not really a problem, and creation of new jobs for data scientists; on the other hand cyberthreats are becoming a serious concern and there are trust issues with data.

1 Introduction

Big data refers to the current trend of massive data acquisition and data utilization that is changing the way knowledge is produced, business conducted, and governance enacted [1]. Specifically, analytics of large pools of data can provide relevant and timely insights to support decision-making processes. While big data has the potential to impact on every sector [2], the problems they face and the uptake of big data are not the same across domains [3]. Moreover, there is a need to enhance the understanding of big data in order to assess the opportunities that can be reaped, e. g. economic gains, but also the risks that the processing of big data can engender, e. g. privacy infringements.

Recent studies have investigated the potential and challenges of big data in different domains, ranging from areas such as health care [4], supply chain management [5], manufacturing [6], smart grids [7], to intrusion detection [8]. Following the line of research in identifying the opportunities and challenges of big data in specific domains, as well as evaluating the uptake of big data together with its societal impacts, this article presents the results of a case study in a domain of high relevance for big data, namely the oil and gas industry.

The oil and gas industry is technically challenging and economically risky [9], requiring large projects and high investments in order to extract oil and gas. The complexity of this industry is reflected by the number of companies associated with the petroleum business – more than 20,000 according to [9]. Some of them correspond to oil operators, large organizations that compete internationally, but also collaborate through joint ventures in order to share project risks. In addition, there is a multitude of vendors that sell equipment and services through the whole oil & gas value chain: drilling, subsurface and top structure (platform) equipment, power generation and transmission, gas processing, utilities, safety, weather forecasting, etc.

With the current crude price crisis situation, many large-scale projects are at risk and oil companies are looking at new ways to improve their margin costs and reduce the uncertainty of their investments – particularly through big data and data science driven solutions [10]. In this context, this article presents the results of an extensive case study in the Norwegian Continental Shelf (NCS) [11]. The aim of our research is twofold: on the one hand, we study the uptake of big data in this industry, analyzing the main data sources employed and their usage; on the other hand, we investigate the societal impacts of big data in oil & gas, studying the economical, legal, social and ethical issues raised by the collection, storage, processing, linking and sharing of data. The case study is focused on the upstream phase of the oil & gas value chain – the most interesting stage in terms of big data [10].

The work on this case study has been primarily motivated by the fact that, in our quest to understand the uptake and impact of big data in the oil and gas industry, we have not found relevant studies in the literature. Core contributions of this case study include identification of big data sources and their current and potential use, economical, social, ethical, and legal impacts of big data, as well as an assessment of the current status of big data in the oil and gas sector. The results of this case study are not only informative, but also provide a timely assessment of the field and motivate additional research in big data relevant to the oil and gas domain.

The rest of the article is structured as follows. Chapter 2 presents the case study and the methodology used to collect relevant information for analysis. Chapter 3 describes the results of the study in terms of identified data sources and their current and potential use in the oil and gas domain, while Chapter 4 focuses on the analysis of results from the perspective of societal impacts of big data in this sector. Chapter 5 provides a current assessment of the extent of big data uptake in this sector.

2 Case Study Overview

In order to accomplish our goals, we considered alternative research designs. However, the exploratory nature of this investigation, its social aspects and the need for in-depth analysis suggest the use of a qualitative method [12]. Moreover, the uptake of big data in the oil and gas industry and its societal impacts is a contemporary phenomenon that requires a detailed contextual analysis of the different stakeholders that intervene. The case study methodology is particularly well suited for this investigation, given that a case study investigates a contemporary phenomenon within its real-life context, when the boundaries between phenomenon and context are not clearly evident, and in which multiple sources of evidence are used [13].

2.1 The Norwegian Continental Shelf

This case study is focused on the use of big data by the oil & gas upstream industry, i. e. exploration and production activities, in the Norwegian Continental Shelf (NCS). The NCS is rich in hydrocarbons that were first discovered in 1969, while commercial production started in the Ekofisk field in 1971 [14].

Petroleum activities in Norway are separated into policy, regulatory and commercial functions: Norway's policy orientation is focused on maintaining control over the oil sector; the Norwegian Petroleum Directorate (NPD) [15] is the regulator body, while petroleum operators compete for oil through a license system. Overall, this separation of concerns is considered the canonical model of good bureaucratic design for a hydrocarbons sector [16].

The oil and gas activities on the NCS repre-

* Guillermo Vega-Gorgojo, Roar Fjellheim, Rajendra Akerkar, Arild Waaler, University of Oslo, Oslo, Norway (E-mail: guiveg@ifi.uio.no); Dumitru Roman, SINTEF / University of Oslo, Oslo, Norway

0179-3187/16/II

sent a large share of the overall industry market, although project complexity is further increased since deposits are offshore in harsh waters and climate conditions are challenging. As a result, petroleum activities in the NCS have prioritized long-term R&D and tackled projects that were highly ambitious technically [16]. This focus on technology makes the NCS especially attractive for a big data case study. Additionally there is a large pool of global operators and oil suppliers, which is representative for the industry. This made it relatively easy to get access to archetypal experts and involve them in the case study.

2.2 Participant organizations

For the realization of this case study we approached various organizations. The profiles of these organizations are described in Table 1, according to the stakeholder taxonomy in [17]. While the industry sector and technology adoption stage dimensions are self-descriptive, the impact of information technology (IT) in industry is an evaluation of an organization's current and future information systems needs [18] that can be classified into four roles:

1 Support role (defensive): current IT is not mission critical for business operations, and new IT will offer little strategic differentiation
2 Factory role (defensive): current IT is critical for the organization's business, but new IT will offer little strategic differentiation
3 Turnaround role (offensive): current IT is not mission critical for business operations, but new IT will be fundamental for the future
4 Strategic role (offensive): current IT is critical for the organization's business, and new IT will be fundamental for the future.

2.3 Data collection activities: interviews and events

We aimed to gather evidence about big data uptake and the societal impacts of big data in the oil & gas industry. The research agenda included big data uptake (data sources, data uses, data challenges, data sharing and open data) and big data societal impacts (economical, social, ethical and legal aspects of data). Multiple sources of evidence were employed in order to augment the validity of our findings, as recommended in [12] and [13]. Interviews were thus arranged with data analysts and IT engineers from the organizations in Table 1.

In all cases some background information were sent beforehand to the interviewees and a brief introduction of the research at the beginning of the interviews was given. The following questionnaire was used that covers the research topics:

Questions about the uptake of big data

– Which are your main data sources?
– What do you use the data for?
– Which are your main data challenges?
– Which are the data sharing practices of your organization?
– Which are your ongoing or future plans for open data?
– Which other parties are involved in your data value chain?
– Which are your main collaborations and data exchanges with other parties?

Questions about societal impacts of big data

– How big data is changing the oil & gas landscape?
– Which are the main legal issues with respect to big data?
– Which are the main risks with respect to policies and legal issues?
– Does your organization consider the social implications of big data Do you use personal or private data in your operation?
– Have you experienced any cyber-attack or suffered data leakage?

We adopted a semi-structured interview approach letting the interviewees speak and elicit their views and opinions, while we aimed to cover all the topics in the agenda and to request further explanations. After each interview we prepared the transcript, had an internal revision and then shared the report with the interviewee (normally within a week), receiving some minor amendments in two cases.

The profiles of the interviewees are shown in Table 2, classified according to the guidelines in [17]. Single interviews of 80 minutes approximately were held with [I- LU], [I-ENI] and [I-SUP]. We were able to interview [I-ST] four times, [I-CP] twice, and [I-NPD-1] and [I-NPD-2] were both interviewed on two occasions at the same time. In these latter cases the sessions were shorter (ranging from 30 to 60 min.), but we had more time available in total. Overall, eleven interviews were conducted for this case study from December 2014 to April 2015.

Besides the interviews, a workshop on big data in oil & gas [19] was held in April 2015. The program included and invited talk from the oil company Statoil, another one from the supplier National Oilwell Varco, a preliminary debriefing of the case study results and a focus group session. We also assisted at a session on big data in oil & gas that was part of the Subsea Valley 2015 conference [20]. All these events were used as input for the case study and rapporteurs had taken notes – Table 3 provides an overview of these additional sources used to collect additional information for analysis.

Throughout this article we profusely include statements from the case study sources – especially in the summary tables, but also within the main text – to support the findings. In all cases the codes included in Tables 2 and 3 were employed to identify the source.

2.4 Analysis of results

Concerning the data analysis, we followed a procedure based on the qualitative [12] and case study [13] research literature. In a first step, the transcripts of the interviews and the additional data sources in Table 3 were prepared. Next, we read through all the transcripts and began the coding of the data – this is the process of organizing the material

Table 1 **Organizations involved in the oil & gas case study**

Organization	Industry sector	Technology adoption stage	Impact of IT in industry
Statoil	Oil & gas operator	Early majority	Strategic role
ConocoPhillips	Oil & gas operator	Early majority	Strategic role
Lundin	Oil & gas operator	Early adopter	Strategic role
Eni Norge	Oil & gas operator	Early majority	Strategic role
SUPPLIER	Oil & gas supplier	Late majority	Turnaround role
NPD	Oil & gas regulator in Norway	Early adopter	Factory role

Table 2 **Interviewees of the oil & gas case study**

Code	Organization	Designation	Knowledge	Position	Interest
I-ST	Statoil	Senior Technical Manager	Very high	Supporter	Very high
I-CP	ConocoPhillips	Data Manager	Very high	Supporter	Very high
I-LU	Lundin	Technical Manager	Very high	Moderate supporter	High
I-ENI	Eni Norge	Technical Manager	Very high	Moderate supporter	High
I-SUP	SUPPLIER	Technical Manager	Very high	Moderate supporter	High
I-NPD-1	NPD	Technical Manager	Very high	Moderate supporter	Medium
I-NPD-2	NPD	Senior Data Manager	Very high	Moderate supporter	Medium

Table 3 **Additional information sources employed in the case study**

Code	Source	Event	Description
FG	seven industry experts in oil & gas from Statoil, National Oilwell Varco, Eni Norge, WesternGeco/ Schlumberger, DNV and Aker Solutions	BYTE oil & gas workshop	Focus group on big data in oil & gas
IT-ST	Knut Sebastian Tungland (Statoil)	BYTE oil & gas workshop	Invited talk: "Big data in subsea – the operator view"
IT-NOV	Hege Kverneland (National Oilwell Varco)	BYTE oil & gas workshop	Invited talk: "Big data in subsea – the supplier view"
T-ST	Sonia Chirico Indrebø (Statoil)	Subsea Valley Conference 2015 – Parallel session on big data	Talk:"Big data in Statoil"
T-McK	Tor Jakob Ramsøy (McKinsey)	Subsea Valley Conference 2015 – Parallel session on big data	Talk: "Digital Energy"

into chunks of text before bringing meaning to information [21]. In particular, we identified specific statements, organized them into categories – corresponding to the topics above – and assigned them labels. Finally, we identified the themes for analysis based on the coding scheme and obtained the findings of the case study.

3 Big Data in the Oil and Gas Domain: Data Sources and their Use

The first result of the analysis resulted in a clear identification of the main big data sources in the oil and gas sector, along with their current and potential use. The sources are described in Chapter 3.1 and their use in Chapter 3.2.

3.1 Main data sources

The interviewees were asked to describe the data sources employed in exploration, operation and production activities. These sources are described below:

Seismic data is the main source for discovering petroleum deposits. Collecting such data is expensive and typically performed by specialized companies such as PGS [22] using seismic vessels that send sound waves deep into the subsurface and a set of hydrophones to detect reflected waves [I-ST]. This process produces significant volumes of data, typically 100s GB per one raw dataset [9]. As explained by [I-ST], operators are obliged to send the seismic data to the Norwegian government – this is incorporated to NPD's Diskos [23], a public database of seismic, well and production data in the NCS (also called Petrobank). Seismic data is also shared among the members of a concession joint venture through Diskos. Since seismic data is a very valuable asset, oil companies take special security measures to conceal it. Seismic surveys are transformed into *3D geology models* – this is probably the most impactful scientific breakthrough of the oil & gas industry [9]. Geologists and petrophysicists analyze these models to find potential deposits of hydrocarbons. Transforming seismic data into 3D models is computing-intensive and results in further amounts of data, 1 TB per one processed dataset. Indeed, Statoil stores around 6 PB of seismic data (raw and processed) [I-ST].

Production data is very important for oil companies and receives a lot of attention. Since this is a commercial-sensitive asset, operators such as Statoil do the accounting of production data by themselves. Oil production is measured at every stage of the flow, while the aggregated figures are reported to the partners in the joint venture and also to the Norwegian Government that has a reporting role.

In the last decade, the oil & gas industry has gone into a process of installing sensors in every piece of equipment – *top-side, subsea and in-well* – and then transferred onshore to a surveillance centre where operations are monitored. New fields are heavily instrumented, e. g. the Edvard Grieg field has approximately 100K data tags [I-LUN-1] and the Goliat field has around 80K data tags [I-ENI]. Sensors are very diverse and generate a lot of data. Moreover, velocity is also a challenge, e. g. a subsea factory produces 100 s of high-speed signals (10 Kbps) and can thus easily generate 1 TB of data per day [I-SUP].

Drilling also generates high-volume and high-velocity data. This data is analyzed in real time for safety reasons and to monitor the drilling process, i. e. to detect if the reservoir is hit [I-ST].

However, integrating the data and presenting in an adequate way to human operators is actually a difficult challenge [I-ST-1, I-ENI-1]. [I-CP] explains that there are some differences on how the data is captured: sometimes the operator has direct access to sensor data, while in other cases, e. g. drilling, the vendor gets the raw data and sends it to the operator. Oil companies also contract services such as vibration monitoring, providing access to sensor data in these cases [I-LU]. Since sensor data is not particularly sensitive, there are more data exchanges among operators and vendors, e. g. for condition-based maintenance of equipment [I-LU].

Document repositories are also quite relevant in the oil & gas industry and employed in different stages. For example, post-drill reports can be analyzed to obtain the rock types in a well – this can be relevant for other analogue areas under exploration. However, document repositories are typically unstructured and quite varied since a report could be produced anytime from the beginning of oil operations in the NCS (1970s). Therefore, the management of knowledge repositories is quite challenging for petroleum companies [I-ST].

Finally, NPD publishes some *reference datasets* as open data – FactPages [24] is one of the most prominent examples, containing information about the activities in the NCS, i. e. licenses, fields, wellbores, discoveries, operators and facilities.

3.2 Usage of data

With such massive data assets collected in the oil & gas industry, there are a number of uses of data in place, as reflected in Table 4 and described in the following paragraphs, organized around the different stages of the upstream value chain.

Exploration and scouting

Seismic processing for the discovery of petroleum is the classical big data problem of the oil & gas industry. Operators have made large investments in high-speed parallel computing and storage infrastructures to generate 3D geology models out of seismic data. The resolution of the images obtained with seismic data is low [9], and for this reason petroleum experts (geophysicists and petrophysicists) try to use additional data sources such as rock types in nearby wells and images from other analogue areas [I-ST]. Nevertheless, the complexity of exploration data makes the access of data to petroleum experts especially challenging, requiring ad hoc querying capabilities. Due to this, the EU-funded Optique project [25] aims to facilitate data access through the use of the Optique platform for a series of case studies, including oil & gas exploration in Statoil [26].

Production

Seismic data is also employed in production for reservoir monitoring, creating 3D models of the reservoir in subsurface. Simulations are then carried out to evaluate how much oil should be produced in a well. Nowadays, there is a trend to permanently deploy seismic sensors in the seabed of a reservoir allowing the detection of microseismic activity. In addition, seismic data from producing fields can be employed to discover oil pockets that can result in more wells for drilling and thus extend the lifetime of a field. Finally, production data is carefully accounted through all stages of the petroleum workflow. Although production data is not especially challenging in terms of big data, it can be combined with other sources to gain further insight, e. g. linking alarms with production data.

Drilling and wells

Drilling operations are normally contracted to specialized companies. Oil operators get the raw data from drilling contractor and then select the target for drilling and decide whether to continue or not, sometimes relying on simulators [I-CP]. These decisions are based on the analysis of drilling data, and they aim to minimize the non-productive time of very costly drilling equipment and crews.
Given the complexity of wells, their integrity is monitored during their complete lifetime. External companies are contracted for well integrity monitoring, employing geological models and using core samples from the well.

Operations

This is possibly the most interesting area in oil & gas in terms of big data [I-ST]. It consists of structured data that is very varied, ranging from 3D models to sensor data. Velocity is also challenging due to the large number of sensors involved producing data in real time. In addition, there are lots of technological opportunities, e. g. Internet of Things. The main drivers for applying big data here include the reduction of well downtime, improving the lifetime of equipment and reducing the number of staff offshore [I-ST].
Among the different uses of data in operations, condition-based maintenance is possibly the one that is receiving more attention. Equipment is instrumented to collect data and analytics are then applied for early detection of potential failures before they occur. Condition-based maintenance is thus much more efficient than traditional reactive or calendar-based approaches. Both operators and suppliers are interested in reducing costs and improving the lifetime of equipment; as a result, there are a number of ongoing collaborations to support condition-based maintenance. Vendors are also analyzing operational data to improve the efficiency of equipment, e. g. using less energy to control the same piece of equipment. The analysis of operational data can also lead to new data-driven products, e. g. Åsgard subsea compressor system [27]. Other opportunities in operations include data-enabled services such as failure detection or vibration monitoring. Integrated operations is another application area that aims to combine data from multiple sources, e. g. operations and production data, and then use analytics to leverage decision-taking processes.

Table 4 **Main uses of data in the oil & gas upstream industry**

Area	Use	Statement [source]
EXPLORATION & SCOUTING	Seismic processing	Seismic processing is the classical big data problem in the oil & gas industry [I-ST] Seismic data is difficult to analyze, complex geo-models are employed [I-CP] Oil companies have made large investments in expensive infrastructures: clusters and high-performance storage [I-ST] New techniques, methods, analytics and tools can be applied to find new discoveries [I-LU]
PRODUCTION	Reservoir monitoring	Seismic shootings are used to create 3D models of the reservoir in subsurface [I-ST] Reservoir simulations are computer intensive and employed to evaluate how much oil should be produced in a well [I-ST] A better understanding of reservoirs, e.g. water flowing, can serve to take better decisions in reaction to events [I-CP]
	Oil exploration	There are also exploration activities in already producing fields to look for oil pockets. This can result in more wells for drilling [I-ST]
	Accounting of production data	Reporting requirements to the authorities and license partners [I-ST-1, I-NPD-1] Not especially interesting in terms of big data by itself [I-ST] Production data can be combined with other data sources, e.g. linking alarms with production data [I-CP]
DRILLING & WELLS	Drilling operations	Drilling data is analyzed to minimize the non-productive time [I-CP] Operators use drilling data to decide whether to continue drilling or not [I-ST]
	Well integrity monitoring	Well integrity monitoring is typically done by specialized companies [I-LU-1, I-ST-1] Geological models are employed, taking into account the type of rock in the well [I-ST]
OPERATIONS	Condition-based maintenance	Equipment suppliers could make better usage of the data, e.g. to optimize equipment performance. Indeed, there is a strong movement towards condition-based maintenance [I-CP] Focus on applying condition-based maintenance [I-SUP-1, I-ST-1, I-LU-1, I-ENI-1, T-ST]
	Equipment improvement	We use operational data to improve the efficiency of equipment [I-SUP]
	Data-driven new products	Some suppliers are using big data to develop new products, e.g. Statoil has expensive equipment that can increase the pressure in a reservoir [I-ST]
	Data-enabled services	Vendors also sell specialized services such as vibration monitoring. For example, SKF is a vendor with expert groups for addressing failures in rotating equipment [I-LU] We are interested in selling a service such as system uptime instead of equipment [I-SUP] Statoil buys services (including data) from the whole supply chain [I-ST]
	Integrated monitoring center	Statoil has a monitoring center for the equipment of each vendor supplier. We are considering replacing them with an integrated center. In this way, it would be possible to get more information from the totality of vendors' equipment [I-ST]
	Integrated operations	Big data can be used for making better and faster decisions in operations by integrating different datasets (drilling, production, etc.) [I-SUP] The analytics of integrated data can be very powerful [I-CP]

4 Societal Impacts of Big Data in the Oil and Gas Sector

The collection, use, sharing and linking of big data implicates a number of economic, social, ethical, and legal issues, including those which may result in positive and negative societal impacts.
Identifying these issues can assist in a better understanding of areas for potential growth and development within the big data industry. This examination also identifies a number of negative societal impacts that arise in relation to the issues addressed. Those negative consequences need to be addressed in order to not overwhelm the potential economic and social benefits of big data.

4.1 Economical impacts

Included in Table 5 are the economical impacts that were found in the oil & gas case study. For each row the specific finding was indicated and a set of statements from the case study data sources that support it.
With the advent of big data in oil & gas, new business models based on data have appeared. One of them is based on data genera-

Table 5 **Economical impacts of big data in the oil & gas upstream industry**

Finding	Statement [source]
Existing business model: Data collection	There are specialized companies, like PGS, that perform seismic shootings [I-ST] Statoil hires other companies for seismic shootings [I-ST]
Data analytics business model	There is a company from Trondheim that has created a database of well-related data (Exprosoft). This company is specialized in projects of well integrity. They gather data from a well and then compare it with their historical dataset using some statistics [I-LU] Wells are more complex and are monitored during their complete lifetime. Well data is processed by an external company [I-ST]
Need for viable data-based business models	Who is paying for the technology? It is necessary to find the business case, since technology-side is possible. The biggest challenge is the business model [IT-NOV] Drilling is a funny business; there are no incentives to drill faster [IT-NOV] There are also economical challenges; we do not have a positive business case for deploying data analytics [FG] How can machine learning companies be players, given the complexity of the oil and gas industry? How can that happen and what will be the effects if that happens? [FG]
Commercial partnerships around data sharing	Condition-based maintenance is an example of an ongoing collaboration with our clients [I-SUP] We have an agreement of 2 years for collaborating with vendors. They will collect data and learn from it, before migrating to condition-based maintenance [I-ENI] We are running pilots for condition-based maintenance; sometimes we do these pilots alone, and other times in collaboration with suppliers. As a result, we have now some equipment in production [I-ST]
Suppliers are actively interested in selling data-driven services	Data-enabled services can be commercialized on top of the equipment sold in order to provide improved services to the clients [I-SUP] Some suppliers want to sell services, not just equipment. This is because they earn more money with services and because they have the experts of the machinery [I-ST] As the manufacturers, suppliers are in the best position to analyze operational data [I-SUP] Suppliers are typically constrained to one "silo", so they are not generally capable of working with big data. Even suppliers like General Electrics (which are good in big data) are limited due to this problem. In contrast, oil companies like ConocoPhillips can provide a holistic view of operations, so they are more naturally capable of doing big data in this area [I-CP]
Open data is a driver for competition	Norway aims to attract investors to compete in the petroleum industry. The FactPages constitutes an easy way to assess available opportunities in the NCS by making openly available production figures, discoveries and licenses [I-NCS-2] NPD began in 1998-1999 to publish open data of the NCS. This is a fantastic way to expose their data and make it available to all interested parties. Before that, companies directly asked NPD for data. NPD has always promoted the openness of data and resources. In this regard, NPD pursues to get as much as possible of the data [I-NCS-1] Companies are also obliged to send the seismic data to the Government – this is incorporated to NPD's Petrobank, i.e. the Diskos database [I-ST]
Companies are somewhat reluctant to open data, but there are emerging initiatives	Statoil is reluctant to share data in exploration, but we have more incentives to share data in operations [I-ST] It could be risky to have access to all the operational data. Exposing commercial sensitive information is a concern for both petroleum operators (in terms of fiscal measures), and for suppliers in terms of equipment and service performance [I-SUP] Some oil operators do not share any data. However, there is an internal debate among operators about this position, and opening data is proposed to exploit added-value services [I-SUP] Operations data is not secret or confidential. We are not very protective as a community [I-LU] Since it is the operator's interest to give access to data to vendors, this is not an issue and access to data is granted [I-LU] There is a problem with different players (driller, operator, reservoir monitor) in the same place, but not sharing anything. How to integrate data that drillers do not have? [IT-NOV]

tion, and one can find companies like PGS that are contracted by petroleum operators to perform seismic shootings. Moreover, datasets such as seismic surveys are traded in all stages of the oil & gas value chain. The data analytics business model is also getting traction: analytics are employed to improve equipment efficiency; some companies are selling specialized services such as well integrity or vibration monitoring; and new products based on data analytics are introduced to the market, e. g. Åsgard compressors.

However, there are some challenges with the business models, requiring funds for investments or other incentives in order to introduce already available new technologies – for instance, [IT-NOV] explained that drilling can be fully automated and thus achieve safer, faster and better drilling; but the business model for deploying the solution is missing. Nevertheless, there are some incipient commercial partnerships around data. For example, petroleum operators and suppliers typically collaborate to apply condition-based maintenance to equipment. Moreover, surveillance centers for monitoring equipment require collaboration among field operators and suppliers – see integrated monitoring center in Table 4.

Given that everybody is realizing the value of data, suppliers are trying to sell data-based services, not just equipment. Since access to data is contract-dependent, this situation creates some tensions. On the one hand, suppliers are in the best position to analyze operational data since they are the manufacturers of the equipment. On the other hand, suppliers are typically constrained to one domain ("silo"), while oil companies are in a better position to provide a holistic view of operations.

NPD closely collaborates with the industry to gather data about petroleum activities in the NCS. This way, NPD aims to promote competition among petroleum operators, embracing open data to facilitate access.

Companies are also considering open data as an opportunity for commercial benefit. Specifically, operators have many incentives to share operations data since privacy concerns are low and there are many opportunities to obtain efficiency gains in operations. However, operators are reluctant to share data in exploration, since it is possible that other parties discover oil deposits. With respect to suppliers, they would prefer to keep the data for themselves, but this is not always possible since data normally belongs to the owner of the equipment (depending on the terms and conditions of the contract). As a result, there are ongoing open data pilots and sharing data collaborations, especially with operations data.

4.2 Social, ethical and legal impacts

Table 6 presents the main social, ethical and legal impacts that were found in the oil & gas case study. In each row the specific finding is indicated as well as a set of statements from the case study data sources that support it.

There is a gap between data scientists and technical petroleum professionals that has not been bridged yet [9]. Nevertheless, the oil & gas industry is becoming interested in hiring data analysts to exploit the potential of big data for the integration of large data volumes, to reduce operating costs and improve recovery rates and to better support decision management.

In this domain, personal privacy is not a big concern and there is little value of social media. Nevertheless, it could be possible to find human errors by analyzing operations data. In contrast, some datasets are highly secret and confidential, so cyber-security measures are quite important and have been adopted through the whole industry.

Traditionally, safety and environment concerns have been pivotal for petroleum activities in the NCS and there are high standards to comply with safety and environment requirements. Big data can help reduce environmental impacts by the early detections of incidents, e. g. oil leakages, and by improving equipment efficiency, e. g. through con-

Table 6 **Social, ethical and legal impacts of big data in the oil & gas upstream industry**

Finding	Statement [source]
There is a real need for data analyst jobs	There are changes in hiring practices, requiring employees with the compe-tences to use the data [FG] There are very few data scientists at ConocoPhillips. We need more [I-CP] Data scientists are not getting into the oil & gas industry. Make a business case and then hire data scientists [T-McK]
Personal privacy is not a big concern	We use industrial data, not Twitter [IT-ST] With big data it could be possible to find who made a bad decision, e.g. a human operator [I-SUP]
There are serious concerns related to cyber-attacks and threats to secret and confidential datasets	Opening up entails some risks. For instance, it could maybe be possible to extract sensitive data such as the daily production of a field [I-SUP] Security/hacking is very much an issue for NPD. Oil & gas in- formation is very important and NPD has a great responsibility. Indeed, companies have to keep trust on NPD. Thus, NPD takes many protective measures such as firewalls and security routines [I-NPD-1] ConocoPhillips has lots of attacks from outside, although we have taken many security measures in IT. Indeed, NPD has instructed oil companies to take measures in this respect [I-CP] The O&G industry is exposed to cyber-threats. Some companies have re-ceived serious attacks; protection measures are needed! [IT- ST]
Big data has the potential to improve safety and environment	Big data can help reduce incidents, e.g. the detection of oil leakages. DTS data can also improve safety when employed for reservoir monitoring [I-CP] Big data helps to give a clear picture of the field operation, and it facilitates the detection of oil leakages or equipment damage [I-SUP] The control system has a lot of alarms and it is literally impossible to manu-ally analyze them all. As an alternative, we can trust the software to auto-matically analyze them [I-CP] I do not see changes due to big data in safety [I-LU] Do we expose the environment for unwanted effects? Statoil wants to know and to show that we don't. We use cameras and sound recorders in the sea (close to the O&G plants), especially if there are big fisheries or corals ne-arby. We want to see if something bad is happening [IT-ST] We are beginning to monitor the seabed before operations. With this data, Statoil can act faster if something is going wrong. We have mobile & fixed equipment capturing video and audio in real time. It can be employed in case of emergency and this data can be shared with others [T-ST]
Concerns exist about trusting data coming from uncontrolled sources	The data ecosystem is complex, and there are many communication ex-changes between oil companies and suppliers – I think that nobody can give a complete overview of the data exchanges in place [I-CP] It is difficult to trust information coming out of the data if you do not have a clear rela-tionship to the underlying reality and if it is not generated by your organization [FG] Those who produce the data only give away aggregated data, and a selecti-on of that aggregated data to specific users. If you want to trust the informa-tion that the system gives you, it can verify that the system is doing what it is supposed to [FG]
Regulation of big data needs clarification	The ownership of operation data is dependent on the contract. Sometimes Statoil can get less data than is captured, while more data could go to supp-liers. This applies to well drilling data and to the machinery on top of a field. This is a complicated ecosystem [I-ST] There is a lot of legislation to take care of. Legislation is different for each country, but there are some commonalities. For example, the data has to be kept at the country of origin, although it is commonly allowed to copy data [I-ST] Legislation of data is still unclear [I-SUP] There is no clear thinking about the regulations with respect to big data yet, and these must be clarified in order to deal with issues around liability, etc. [FG] Making raw data regulated is something that has to be judged on the critica-lity of the risk. Ideas like black boxes could carry over into this industry be-cause the risks of malfunction can be so severe[FG]

dition-based maintenance. There are also pilot initiatives that can be highly valuable to assess the impact of oil extraction activities and to act faster in case of an accident:

- Statoil is deploying myriads of sensors at the sea bottom to monitor reservoirs. For a high-resolution image the microphones need to be in the same place and for this reason they are placed on a permanent basis in the seabed.
- Seismic shootings are taken each six months, but it takes months to get the processed data. This data can feed a simulator and the results used to decide to drill a new well, extract more oil and gas, or inject water to keep the pressure up – if the right decisions are taken, recovery rates of the reservoir can be significantly improved.
- However, it is possible to do more with the cables and sensors in the seabed. Indeed, Statoil is collecting data every second to detect microfractions. This signal is used to decide whether to increase or not the pressure in the reservoir, resulting in better recovery rates. Environmentally this is also good, since Statoil can use the sensors to detect oil leakages. [IT-ST, T-ST]

There is also a trust issue with data coming from uncontrolled sources. This is especially relevant when aggregating data or when applying data-driven models.

With respect to legal impacts, petroleum activities in Norway rely on a mature regulation framework that enforces the separation of policy, regulatory and commercial functions. The Petroleum Act No. 72 of 29 November 1996 [28] provides the general legal basis for the licensing that governs Norwegian petroleum activities. This is the result of many years of close collaboration of NPD with field operators. These have reporting obligations for seismic and production data, but receive support on legislation about safety, licensing and other issues. As a result, all players have trust in NPD and accept their obligations in the petroleum industry.

While production and seismic data are highly regulated by the authorities, other datasets, e. g. operations data, are normally regulated by the terms and conditions of a contract. In this regard, the owner of data is normally the owner of the equipment that produces the data. There are some exceptions, though – for instance, drilling companies normally collect the raw data that is then supplied to operators. Furthermore, international legislation is problematic for oil companies, since different laws apply to each country. As a result, legislation of big data aspects requires additional clarification.

Indeed, industry stakeholders are becoming increasingly aware of the value of data. This is quite evident for seismic data that is commonly traded or auctioned. For other datasets, ownership of data will possibly be subject of contention.

5 Discussions on the Uptake of Big Data in the Oil and Gas Sector

In our fieldwork we collected a number of testimonials, impressions and opinions about the adoption and challenges of big data in the oil & gas industry. With this input Table 7 was elaborated, containing the main insights and the statements that support them.

The assessment reveals that the oil & gas industry is beginning to adopt big data: stakeholders are collecting as much data as possible, although there is some criticism about its actual usage in practice – this suggests an awareness of the potential of big data in oil & gas.

While this industry is quite familiar with high volumes of data, exponential growths can be expected in the near future, as new devices to track equipment and personnel performance are deployed everywhere and collecting more data than ever. Nevertheless, volume is not the only data challenge that the oil & gas industry is facing; variety and velocity are becoming increasingly important as more data signals are combined and analyzed in real-time. Moreover, humans cannot deal with such amounts of data, requiring effective tools for visualizing, querying and summarizing data.

Big data advocates propose to find correlations and patterns in the data, without requiring a preliminary hypothesis – this is sometimes referenced as "let the data speak" [29]. In contrast, the oil & gas industry relies on

Table 7 **Assessment of big data in the oil & gas upstream industry**

Insight	Statement [source]
Big data in oil & gas is in the early-middle stages of development	Big data is still an emerging field and it has not yet changed the game in the oil & gas industry. This industry is a late adopter of big data [I-CP] Everybody is talking about big data, but this industry is fooling around and doing small data [T-McK] Big data is quite new for SUPPLIER [I-SUP] This industry is good at storing data, but not so much at making use out of it [I-CP] Oil and gas is still at the first stage of big data in the sense that it is being used externally but not to acquire knowledge for themselves. For example, lots of data about what happens when the drill gets stuck, but they are not using that data to predict the drill getting stuck. Structured data plus interpretation/models are not being converted into knowledge [FG] There are a lot of areas that can be helped by big data. How can we plan when to have a boat coming with a new set of pipes? [FG] Machine learning is beginning to be integrated into technical systems [FG]
More data is being made available in the oil and gas field	In exploration, more sensors are employed, and microphones for collecting seismic data are permanently deployed at the seabed in some cases [I-ST] ConocoPhillips has hundreds of TBs from the Ekofisk area. Volume is an issue, since seismic datasets are growing [I-CP] PRM (Permanent Reservoir Monitoring) will push volume of seismic data from the Terabyte to the Petabyte region, due to more frequent data collection [I-CP] Statoil has 8PB of data and 6PB are seismic. Seismic data are not structured and are stored in files [I-ST] The volume of sensor data is big (TBs and increasing), with little metadata [I-ST]
Variety and velocity are becoming important challenges	Operations data is very varied, ranging from 3D models to sensor data, and velocity is also a challenge [I-ST] Any piece of equipment is identified with a tag, e.g. pipes, sensors, transmitters. On Edvard Grieg field there are approx. 100.000 tags. Lundin has 10K unique instruments, each collecting approx. 30 different parameters on the average [I-LU] Scouting for hydrocarbons involves a huge analytical work in which the main challenges are volume, quality and, especially, variety [I-ST] A subsea factory is very advanced equipment consisting of several connected processing components. It can generate 100s of high- speed signals (~10Kbps). Thus, it can easily generate 1 TB of data per day. It will typically use optical fiber connection with high bandwidth [I-SUP]
Data overflow and visualization of data	In the Macondo blowout in 2010 there was so much data that operators could not take an action in time. As humans we cannot deal with all the data [IT-NOV] In operations the visualization of data is not sufficiently effective and comprehensible. Something is missing with respect to the user, even if you have a monitor, you need to interpret what is presented and the interconnections of data are not evident [I-ENI] There are lots of data coming in from different components. A challenge for the operator is how to pay attention to/align the information coming in on 15 different screens. How to simplify this into manageable outputs? [FG]
Analytics with physical models VS data-driven models	An important question is how to do analytics. One classical way is to employ physical models. Another path is just looking for correlations [I-CP] We normally employ physical models, while another possibility is the use of data-driven models – although their value has to be proven here. Statoil is currently trying different models with the available data [I-ST] In some sectors there is the idea that you should "let the data speak for itself" but in the more classical oil and gas approach, you will base the analytical models on equations and models (physics) [FG] We have tested the distinction between the physical models and the machine learning models. Two years ago, the physical models performed better, but the machine learning models are constantly evolving [FG]
Resistance to change	A lot of the technology is there, but the mindset is the main problem [IT-NOV] It is extremely difficult to change the drilling ecosystem because of the different players involved – many of them are reluctant to introduce changes [I-ST] There are many possibilities to reduce production losses by analyzing the data, but the business side is not ready yet to look into this [I-CP]
Effectiveness of big data in oil & gas	Everybody is trying to do big data, but the industry needs success stories to know what can be really done with big data. Right now, it is not easy to foresee what can be done; there are some analytics and time series analysis under way, but next level is to get real knowledge out of the data [I-SUP] Big data analytics introduces uncertainty, but we don't have so much experience with big data so as to report concerns [I-CP] It costs something to analyze 2000 data points, and you have to have a good reason to invest in that analysis [FG]

well-established physical models for doing analytics. This disjunctive between physical and data-driven models is currently under discussion in this domain.

Still, there is some resistance to embrace big data practices and techniques in oil & gas. In many cases the technology is already available, but decision-takers are somewhat reluctant to introduce changes – especially if business models are affected. Nonetheless, the effectiveness of big data has to be proved in oil & gas, and the industry needs success stories that showcase the benefits that can be reaped.

6 Conclusion

The oil & gas domain is transitioning to a data-centric industry. While big data still needs to prove its effectiveness in oil & gas, the industry is beginning to realize its potential and there are many ongoing initiatives, especially in operations.

In this case study a number of economical impacts associated with the use of big data in oil & gas were identified: data generation and data analytics business models are beginning to get traction, there is a number of commercial partnerships around data and the Norwegian regulator has embraced open data in order to spur competition among oil operators. However, companies are still reluctant to share their data, despite some emerging initiatives. Moreover, existing business models have to be reworked in order to promote the adoption of big data.

On the positive side of social and ethical impacts, safety and environment concerns can be mitigated with big data, personal privacy is of low concern in oil & gas and jobs are created for data scientists – whereas operators and other types of jobs might be less in demand. On the negative side, cyber-security is becoming a serious concern and there are trust issues with third-party data and data-driven analytics.

The petroleum industry benefits from a mature regulation framework in Norway, although regulation of data requires further clarification. Additionally, companies are increasingly aware of the value of data and we can expect contention about data ownership. Many companies in the oil business are multinationals, so there is a need to harmonize international legislation with respect to data. Indeed, some vendors are becoming leaders in big data, and the rest should embrace big data in order to succeed in the future.

This research has been partially funded by the European Commission through the BYTE (FP7 GA 619551) project. The authors would like to thank the interviewees and organizations in the Norwegian Continental Shelf who participated in this case study. The views expressed by the interviewees are their own and do not necessarily represent the views of their employers.

References

[1] Kitchin, R.: The data revolution: Big data, open data, data infrastructures and their consequences. Sage (2014).

[2] Probst, L., Monfardini, E., Frideres, L., Clarke, S., Demetri, D., Schnabel, L., Kauffmann, A.: Big data: Analytics & decision making. Business innovation observatory, European Union (2013).

[3] Chen, C. P., Zhang, C. Y.: Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences 275 (2014) 314–347.

[4] Murdoch, T. B., Detsky, A. S.: The inevitable application of big data to health care. Jama 309(13) (2013) 1351–1352.

[5] Schoenherr, T., Speier-Pero, C.: Data science, predictive analytics, and big data in supply chain management: Current state and future potential.

Journal of Business Logistics 36(1) (2015) 120–132.

[6] Bi, Z., Cochran, D.: Big data analytics with applications. Journal of Management Analytics 1(4) (2014) 249–265.

[7] Corbett, J., Webster, J.: Organizational sensemaking and big data frames: Opportunity, control, and data limitation. In: System Sciences (HICSS), 2015 48th Hawaii International Conference on. (2015) 4772–4781.

[8] Zuech, R., Khoshgoftaar, T. M., Wald, R.: Intrusion detection and big heterogeneous data: a survey. Journal of Big Data 2(1) (2015) 1–41.

[9] Farris, A.: How big data is changing the oil & gas industry. Analytics Magazine (2012) 20–27.

[10] Cowles, D.: Oil, Gas, and Data: High-Performance Data Tools in the Production of Industrial Power. O'Reilly (2015).

[11] Norwegian Petroleum Directorate: Norwegian continental shelf – questions and answers. https://www.regjeringen.no/no/dokumenter/the-continental-shelf – questions-and-an/id583774/ (2016); last accessed March 2016.

[12] Creswell, J. W.: Research design: Qualitative, quantitative, and mixed methods approaches. Sage Publications (2009) 3rd edition.

[13] Yin, R. K.: Case study research. Sage Publications (2014) 5th edition.

[14] Tormodsgard, Y.: Facts 2014. the Norwegian petroleum sector. Technical report, Norwegian Ministry of Petroleum and Energy and Norwegian Petroleum Directorate (2014).

[15] Norwegian Petroleum Directorate: Norwegian petroleum directorate website. http: //npd.no/en/ (2016) Last accessed March 2016.

[16] Thurber, M. C., Istad, B. T.: Norway's evolving champion: Statoil and the politics of state enterprise. Program on Energy and Sustainable Development 92, Stanford University (2010).

[17] Curry, E., Freitas, A., Vega-Gorgojo, G., Bigagli, L., Løvoll, G., Finn, R.: Stakeholder taxonomy. Deliverable 8.1, BYTE project (2014).

[18] Nolan, R., McFarlan, F. W.: Information technology and the board of directors. Harvard business review 83(10) (2005) 96–106.

[19] BYTE Project: Big data in oil & gas workshop. http://byte-project.eu/wp-content/uploads/2015/05/BYTE-Oslo-events-agenda-April-2015.pdf (2015); last accessed March 2016.

[20] Subsea Valley: Programme of the subsea valley conference 2015. http://subseavalley.com/nyheter/arkiv/2015/01/konferanseprogrammet/ (2015) Last accessed March 2016.

[21] Rossman, G., Rallis, S. F.: Learning in the field: An introduction to qualitative research. Sage, Thousand Oaks, CA, USA (1998).

[22] Petroleum Geo-Services (PGS): PGS website. http://www.pgs.com/ (2016); last accessed March 2016.

[23] Norwegian Petroleum Directorate: Diskos website. http://www.npd.no/en/about-us/collaboration-projects/diskos/ (2015); last accessed March 2016.

[24] Norwegian Petroleum Directorate: FactPages website. http://factpages.npd.no/factpages/ (2016); last accessed March 2016.

[25] Optique Project: Optique project website. http://optique-project.eu/ (2016); last accessed March 2016.

[26] Giese, M., Soylu, A., Vega-Gorgojo, G., Waaler, A., et al.: Optique – zooming in on big data access. IEEE Computer 48(3) (2015) 60–67.

[27] Aker Solutions: Åsgard subsea gas compression system. http://citrix.akersolutions.com/en/Global-menu/Projects/Subsea-technologies-and-services/System-engineering/Aasgard-subsea-gas-compression/(2016); last accessed March 2016.

[28] Norwegian Petroleum Directorate: Petroleum act no. 72 of 29 November 1996. http://www.npd.no/en/Regulations/Acts/Petroleum-activities-act/ (1996); last accessed March 2016.

[29] Mayer-Schönberger, V., Cukier, K.: Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt (2013)

Guillermo Vega-Gorgojo *is a postdoctoral researcher in the Department of Informatics at the University of Oslo in Norway. His research interests include big data, semantic web and visual search interfaces. He has authored more than 60 scientific publications. Guillermo holds an associate professorship at the University of Valladolid, Spain.*

Roar Fjellheim *is Director of Business Development at Computas, a Norwegian IT company delivering services and solutions to customers in the oil & gas industry and other domains. Fjellheim received his education in control engineering and computer science at NTNU (Trondheim, Norway) and has 40 years of experience from industry and academia. He is also an adjunct professor at the University of Oslo (informatics).*

Dumitru Roman *works as a Senior Research Scientist at SINTEF ICT (Norway). He has wide experience with initiating, leading, and carrying out (research-intensive) projects on data management and service-oriented topics, being involved with over 20 large-scale R&D projects at the European level during the past 12 years. He is currently active in the data management field, particularly in the emerging Data-as-a-Service (DaaS) domain. He holds an adjunct associate professorship at the University of Oslo, Norway.*

Rajendra Akerkar *has adjunct position in University of Oslo, Norway. He is a senior researcher at Western Norway Research Institute, Sogndal. His recent research focuses on application of big data technologies to real-world challenges, and social media analytics in a wide set of semantic dimensions. His research also addresses the social impacts of big data. He has authored/co-authored 14 books, 124 research articles, co-edited 2 e-Books and edited 14 volumes of international conferences and workshops.*

Arild Waaler *is a professor in the Department of Informatics at the University of Oslo where he established a research group that targets semantics-based approaches to data integration and data access, with special focus on applications to the oil and gas sector. He currently leads the Optique project, a large-scale EU funded initiative, and is currently serving as director of the SIRIUS Centre for Research Based Innovation.*