Analyzing the Roles and Competence Demand for Digitalization in the Oil and Gas 4.0 Era

Konstantinos Georgiou¹, Nikolaos Mittas², Ioannis Mamalikidis¹, Athanasios Mitropoulos² and Lefteris Angelis¹

¹ School of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
² Department of Chemistry, International Hellenic University, Ag. Loukas, 65404 Kavala, Greece

Corresponding author: Nikolaos Mittas (e-mail: nmittas@chem.ihu.gr).

ABSTRACT The rapidly moving technological advances have forced key industrial sectors to accelerate their digital transformation processes. Despite the growing interest of organizations on investing in technological advances, the Oil and Gas (O&G) industry is losing ground on the digitalization race. Except from the technological inhibitors, the relatively slow transition is related to human capital and the lack of skillful workforce able to fulfill emerging digital job positions. Additionally, the empirical evidence indicates a knowledge gap regarding the current demand for O&G digital experts and the associated responsibilities and competencies. The objective of this study is two-fold: First, we investigate the status of the O&G labor market digital career opportunities. Second, we shed light on the most prominent digital job roles and the corresponding skillsets, so as to design competence maps with interrelated technological skills. To meet our objectives, we analyzed 1999 job openings from three well-known O&G job boards through a framework augmented with natural language processing and graph theory community detection methodologies. The findings for emerging technology trends constitute an empirical benchmark for a wide range of target groups that are interested in acquiring knowledge related to the state-of-the-art digital demand and the endorsed competencies in the uprising O&G digitalized era.

INDEX TERMS Oil & Gas 4.0, labor market analytics, competence mining, digital skills

I. INTRODUCTION
The fast-moving digital innovations, which are significant drivers of the fourth industrial revolution (or Industry 4.0), has forced most of the key industries to initiate a transition journey on their business processes and models [1]. Although there is not a global definition for Digital Transformation, the term is used for representing the strategy of adopting cutting-edge Information and Communication Technologies (ICT), such as Big Data Analytics, Internet of Things, Cloud services etc. with the aim of reshaping, evolving or replacing traditional processes across all aspects of an industry sector at both intra- and inter-organizational levels.

The Oil & Gas (O&G) industry could not be an impassive actor to this revolution marathon, since it is considered one of the biggest industrial sectors in terms of dollar value [2]. Nowadays, the O&G sector is under unceasing supply and demand pressures caused by the recent decrease in prices and production [3,4], the shift in green and renewable energy solutions [5-8], the imperative need for reduction of the operating expenses [9-11] and the fragile geopolitics in regions with a critical role in the global energy transition [12]. Moreover, the Covid-19 pandemic crisis has forced even the most traditional industries to seek for innovative ways to transform their services and explore new opportunities [13-15].

Based on findings [16], the digital transition into the new “O&G 4.0” era [17] is estimated to “unlock approximately $1 trillion of value for the industry and $640 billion for its customers and wider society” during the decade 2016-2025. To achieve this, stakeholders confront significant organizational and operational barriers realizing that digital technology should become a strategic priority [17]. The importance of digitalization and the incorporation of emerging ICT solutions in the O&G sector is also highlighted in a study¹ launched by Deloitte² [18], stating that the O&G industry can potentially increase its productivity and reduce costs in its operations, while generating a large profit margin. The Digital Operations Transformation (DOT) model proposed by the

¹https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Energy-and-Resources/gx-online-from-bytes-to-barrels.pdf
²https://www.deloitte.com
company is a full-fledged framework that sets necessary guidelines for an O&G company in order to digitalize its services without difficulties.

Despite the great promises for business value creation through the adoption of digital technologies [16], the transition into O&G 4.0 is certainly not a bed of roses, due to significant inhibitors that need to be recognized, analyzed and overcome. The decelerating factors can be categorized into two main pillars: (i) the technological challenges related to the adoption of emerging trends and (ii) non-technological factors such as the lack of standardization frameworks in the collection process of complex and unstructured data streams from multiple digital sources [16], the absence of integrated platforms across the value chain augmented with regulatory frameworks for data sharing and cybersecurity threats [19], as the most significant technological challenges.

Besides the technological barriers, the digital transformation is also hindered by other non-technological aspects related to both organizational structure and human capital. Generally, resistance to change constitutes a cultural obstacle that each organization must face in all levels of the management hierarchy [20]. In particular, at the upper levels of the organizational structure, the root of the problem is associated to the people-centric management model followed by the majority of O&G firms and the lack of inspired leaders with the corresponding digital skills, competencies and experience who overlook the created value and benefits of the digital transformation [17]. Moreover, traditional leaders are usually unwilling to accept and adopt digital solutions in their business processes, since they fear taking risks in the light of catastrophic economic impact on their organizations [21-23]. Moving down to the middle levels of human resource hierarchy, despite the increasing interest of open-minded leaders to embrace the new wave of change through the adoption of ICT, the lack of talented and skilled digital workforce is among the key factors slowing down the application of digital technologies in the O&G sector [24].

To overcome the abovementioned barriers, the digital experts synthesize a roadmap with general recommendations and guidelines. First of all, the policymakers have to clearly realize that a digital culture following the state-of-the-art innovations of technology must become an organizational priority [25-28]. In addition, the leaders should speed up the transition by investing in human capital that will reinforce the new way of digital thinking [29,30]. Moreover, the digital transformation is a life-long learning process and for this reason, leaders should empower and reinforce the learning curve by developing digital educational programs [31-33].

Finally, higher educational institutions have to be confronted to the current needs of Industry 4.0, since there is noted a significant lag regarding the adoption of the recent technological advances into the existing educational programmes [34].

Hence, the arising critical question is whether the O&G industry has been aligned to the current trends of digital transformation. The findings are far from disappointing highlighting that the O&G industry with the lowest average score on Deloitte’s digital maturity index, brings up the rear compared to other fifteen sectors. The report concluded that O&G stakeholders should put all their effort on the establishment of an organizational culture motivated by digital leadership and talented digital workers that will develop the roadmap for the next generation workforce.

Although the empirical evidence reveals that stakeholders agree on the fact that the transition into O&G 4.0 should become a continuously evolving task of extremely high priority, it seems that there is a significant gap in the body of knowledge related to the current demand for the next generation of digital workforce in this specific industry. More specifically, the main body of existing literature research focuses on the digitalized concepts and technological innovations synthesizing the new era of O&G 4.0 and how these could be adapted to the special needs of the O&G industry, whereas other studies discuss the opportunities, barriers and other organizational challenges related to this digital transformation process [35,36].

Based on these considerations, the motivating idea behind this paper is the investigation of a crucial aspect of the digital transformation journey referring to the investment on human capital in O&G 4.0. To meet this goal, in this study, we aim at: (a) systematically investigating the current demand for digital workforce and (b) identifying the associated (i) job roles and (ii) required competencies for fulfilling a job position in the O&G 4.0 industry. Apart from the extraction of the most prominent competencies, the proposed approach provides a mechanism for the exploration of interconnected cutting-edge technologies serving, in turn, to the design of competence maps within emerging technology trends. These competence maps define, in fact, a combination of key skills (or skillsets) within each technology field providing a more comprehensive view rather than an isolated list of dominant competencies concerning the requirements for fulfilling digital job positions. To meet our objectives and provide answers to certain research questions, the proposed framework is based solely on mining knowledge from well-known O&G job boards and more precisely, on the application of natural language processing and text mining methodologies on data extracted from relevant job postings. Furthermore, advanced statistical methods that are based on graph theory and community detection

---

3https://www2.deloitte.com/gz/en/pages/about-deloitte/articles/disrupt-the-norm/standing-still-not-option.html
algorithms are deployed for synthesizing competence maps of talented digital workforce.

The proposed approach offers an alternative perspective to other traditional approaches for analysing technology trends and forming of competence maps that are based on qualitative methods, such as surveys, case studies, focus groups and interviews with senior executives and project managers. In addition, the proposed approach is able to capture the rapid changes of digital technologies and their associated competencies and thus, it constitutes a valuable tool for a wide range of competence management purposes. To the best of our knowledge, this is the first research attempt shedding light into aspects related to human capital and the required digital competencies in today’s gradual transformation journey of the O&G 4.0 industry. We believe that the outcomes of this study contribute to the scientific body of knowledge, but they also offer practical insights and implications that can be useful for a wide range of target groups including senior executives and human resource managers, higher educational institutions, students and job seekers and generally, society.

The rest of the paper is organized as follows: In Section 2, we present background information related to underlying concepts of the study. In Section 3, we present the research objectives and the research questions posed in this study. In Section 4, we present in detail, the proposed approach and its corresponding phases, while in Section 5, we present and discuss the results of the analysis conducted on the experimental setup. In Section 6, we discuss possible threats to validity of our approach. Finally, in Section 7, we conclude the paper with a discussion concerning the implications of the study to stakeholders and target groups and directions for future work.

II. BACKGROUND INFORMATION

In this section, we present background information necessary for facilitating the understanding of the current study. Our focus on presenting related literature was in tracking the recent rising interest in the digitalization of the O&G industry, digital initiatives in Industry 4.0 and their impact on new opportunities related to digital workforce, as well as the utilization of Competence Management for the examination of current digital demand and the recruitment of appropriate candidates.

A. INDUSTRY 4.0

Nowadays, Industry 4.0 has arisen as the new face for the majority of industry sectors and the transformation of traditionally physical services to digitalized solutions that do not require human intervention. It is a constantly developing idea encapsulating several individual domains that rely on modern technological feats to operate. The research interest is continuous and evolving [37], with primary focus given in the concept and content of Industry 4.0, the main technologies required for its implementation along with its potential applications.

At its core, Industry 4.0 takes advantage of specific technological innovations such as the Internet of Things (IoT) and big Data Analytics (DA) in combination with cyber-security systems to achieve communication between machines and frameworks in order to automate a wide range of industrial processes [38]. It has been characterized as the “Fourth Industrial Revolution” and it is expected to create job opportunities for talented scientists and workforce, while retaining its profitable rise [39]. It has gained increased support in countries such as Germany and China, with governments forming long term plans lasting at least until 2025 in order to integrate Industry 4.0 to their industrial manpower [1]. The application field is not limited to operational and business processes, but it can be extended to the production of smart products, home applications equipped with IoT frameworks, urban and transport services and the improvement of facilities [37].

The key concepts of Industry 4.0, apart from leveraging smart solutions and technologies, involve the growing need for flexibility and efficiency in the industrial market. Having in mind the exponential increase in data production, industrial life cycles have acquired a tendency of moving towards digitalization and datafication [40], with brief development cycles and automated processes [41]. In parallel, efforts are being made in transforming businesses from corporations that simply serve a demand to socially accepted organizations, tailored to the public’s needs. Thus, the big challenge of Industry 4.0 is the detection of innovative methodologies and frameworks that do not only cater to large-scaled, data-driven projects, but also to be constantly aware of the rapidly changing community standards.

A crucial factor for the transition to Industry 4.0 is the notion of sustainable development [42], in the sense that all departments of an organization should be interconnected and operate as a whole entity. An alarming concern for the higher productive paces of businesses that adhere to the Industry 4.0 model, is the environmental ramifications and the uncontrolled energy consumption. However, research has indicated [43,44] that adopting circular economy protocols and focusing on sustainable management and logistics can potentially decrease the energy footprints of the industry while maintaining efficient production rates. The IoT and Cloud technologies are key players in this effort to respect the environment and develop sustainable business models, as their applications can easily moderate business activity and handle data accordingly. Not limited to environmental drawbacks, challenges for a smooth adoption of Industry 4.0 models to the
industrial market are limitations in data processing power, the threat of cyber-attacks indicating the imperative need for secure cyber systems and the overall hesitation of the public towards new technologies that leads to investment boundaries [45].

B. COMPETENCE MANAGEMENT
Although the concept of Industry 4.0 is expected to bring benefits to many organization levels of companies, it also hinders many challenges related to the establishment of an innovative culture through the investment in digitalized human capital. More specifically, the reliance on digital services creates a growing demand for talented personnel, able to handle this industrial evolution. Thus, dedicated processes for competence management and assessment are necessary in order to select appropriate individuals to fill digital-based job positions. However, a common mistake that job recruiters are prone to when seeking suitable candidates for a specific vacancy is that they do not assess the particular skillsets necessary for the position. This leads to positions that do not respond to the current skills of professionals, technological or otherwise, thus creating a gap between the recruiter and the applicant.

After the digitalization of the industry, many platforms have been actively promoting the e-recruitment strategy by posting jobs online and constantly enhancing their descriptions via exploiting job-related data. A significant categorization of competencies, that facilitates the search for an appropriate candidate is the separation in soft skills and hard skills. In essence, soft skills concern the interpersonal characteristics of an individual, not linked to technical prowess but linked to his/her communication and collaboration spirit. In contrast, hard skills directly comprise the technical knowledge of an individual in a scientific field and typically measure his/her suitability for a position. Both types of skills complement each other and shape the profile of a job candidate.

Competency Management (CM) [46-48] can be leveraged as an efficient way of recognizing the specific needs for a job position and assign it to the proper professional, thus refraining from wasting resources and time. It is a concept that can support a wide range of strategic decisions aiming to the prediction of demand for current and future competencies. By exploiting its potential, companies can not only design training programmes for their personnel but also develop improved career prospects and performance indicators for their workforce. In parallel, they can have a constant source of feedback for their decisions and labor management [49]. To achieve this enticing prospect, modern organizations, and particularly sectors that desire to follow the Industry 4.0 guidelines, adhere to e-recruitment strategies and adaptation of job analytics processes in order to both attract and hire professionals and monitor their existing employees.

For these reasons, job postings provide useful insights about the current state of labor market digital needs, the requirements of a specific sector or the interest of candidates regarding different types of jobs. Summarizing, the job analytics research domain focuses on the collection of data related to e-recruitment phases, online job postings and the soft/hard skills associated with the labor market. The subsequent analysis of such data to extract useful information is a growing trend that reveals several aspects of the workforce.

C. O&G INDUSTRY 4.0
The O&G industry, as stated in the Introduction, is facing difficulties in transitioning to the digitalization of its upstream, midstream or downstream processes. However, as the accumulated data gathered from upstream procedures are of considerable volume and importance, the adoption of (big) DA technological advances have been proposed as early as 2015, when the term “Big Data” initially gained attraction and showcased the functionality of exploiting large amounts of data for industrial and research purposes in this specific sector. More specifically, it is implied that the O&G industry should strive to revolutionize its data processing policies and opt to imitate other large industries (healthcare, financial, retail etc.) in uncovering and predicting more complex relationships between extracted data [54]. The importance of Big Data is supported by their functionality in knowledge extraction and their contribution to the fusion of information via the knowledge pyramid [55]. Thus, O&G industries should give primary focus in accessing available data and exploit their industrial potential for knowledge creation and management.

Mohammadpoor and Torabi [56] note the importance in adoption of (big) DA as a rising trend in O&G and highlight the multiple categories, in which they can be leveraged to maximize potential. As for upstream services, they initially showcase the possibility of exploiting machine learning methodologies to detect patterns in seismic activity and petroleum exploration in order to predict unusual or more rough geological intricacies and adjust the drilling processes. The reservoir and production engineering processes can also be benefited by the analysis of DA in order to design more refined performance or transportation plans, study the chemical behavior of oil and gas and improve the miscellaneous equipment utilized during the extraction phases (drills, pumps etc.). However, the data quality and the advanced complexity for finding a solution to these problems are listed as challenges demanding immediate attention.

Hence, DA advances can be considered as the core ingredient for transferring the operational prowess of O&G through the technological spectrum, recent literature suggests the imperative need of designing data processing methods with respect to privacy and security [19]. As data gathered from upstream O&G processes are very important for supporting various society infrastructures and the eventual downstream transformation of petroleum to useful products, a potential breach of security or a misalignment in communication between sensors, frameworks or machines could lead to catastrophic financial and practical
consequences. Thus, the exploitation of Cloud and IoT technologies is an obvious solution for data protection and the creation of digital solutions that can operate with increased amounts of information. Further focus in data privacy reveals that large-scale sensors and frameworks, utilized in drilling plants are highly susceptible and vulnerable to cyber-attacks, and should be well designed to avoid a possible data leakage [36]. Thus, Cloud and IoT technologies should be employed to secure scalability and transparency. Similar consensus is reached in another study [17], which encompasses DA, IoT and its industrial counterpart (Industrial Internet of Thing (IIoT)) and Cloud solutions under an umbrella term (Oil & Gas 4.0). Moreover, it emphasizes the imperative need for the sector to answer the continuous changes in the industrial landscape and exploit the overabundance of data produced by its activities. To achieve this objective, the study makes clear the obstacles that have to be overtaken, ranging from governmental setbacks to the reduced cultivation of multidisciplinary job fields, but also presents the engaging opportunities in production efficiency, lifecycle management and energy consumption.

A driving factor for these solutions to be gradually implemented and inserted to the industrial routines of O&G is the creation of employment opportunities and job vacancies, as indicated by research literature that not only clarifies the importance of data but also the requirement for specialists that can comprehend and analyze them [54,56,57]. A recent study [57] moves the employment field one step further by defining high ranking positions necessary in a modern O&G industry for efficient decision making, such as a Chief Information Officer. However, the existence of Data Scientists, Big Data Experts and all subsequent variations of these job roles should be of primary concern for companies. In addition, DA, IoT and Cloud technologies are praised as the torchbearers of innovation in the O&G industry by the World Economy Forum [17]. Thus, their integration to business models should be a prerequisite when designing and deploying new services and solutions.

III. RESEARCH OBJECTIVES & QUESTIONS

The main goal of this study according to the Goal-Question-Metric (GQM) approach [58] can be defined as follows: Analyse the demand of digital workforce, for the purpose of identifying the characteristics, job roles and skillsets from the point of view of industry practitioners and other stakeholders in the context of the O&G 4.0 Industry. To meet this goal, we formulated the following research questions (RQs) that can be conceptualized into two specific pillars as presented in the introductory section: (a) investigation of the current state of demand for digital workforce and (b) identification of (i) job roles driving the digitalization process and (ii) competencies accumulated, in turn, to specific job profiles (skillsets) necessary for fulfilling job positions in the O&G 4.0 industry.

[RQ1] What is the current demand for digital workforce in the O&G 4.0 era?
Motivation: As we have already mentioned, despite the fact that key industry sectors have radically reshaped their organizational structure and processes by adopting emerging digital technologies, the O&G industry is the straggler to this transitional marathon. Given the imperative need for understanding this phenomenon, the intention of RQ1 is to provide an overview of the current needs for high-skilled digital jobs. Generally, our interest focuses on the distribution of job openings across emerging technology trends, the geographical distribution of the digital workforce demand and the identification of inspiring companies driving the transition into the O&G 4.0 era. To facilitate the examination of RQ1, we state three relevant research sub-questions:

- [RQ1.1] Which are the characteristics of job offers in terms of technology trends, employment type and year of job opening?
- [RQ1.2] How are these job offers distributed at geographical level across the technology trends?
- [RQ1.3] Who are the leading players (companies) driving the digital transition into the O&G 4.0 era?

[RQ2] Which are the digital job roles with the highest demand in the O&G 4.0 era?
Motivation: In line with the recent technological advances, senior executives of the O&G companies have realized the necessity of attracting and recruiting digital specialists and/or develop a vision of digital mindset through on-the-job training to their current employees. However, the term digital transformation is, in fact, a broad concept encompassing a wide range of technological initiatives and solutions across multiple aspects of the O&G industry. Due to this fact, the assignment of “the right person with the right expertise to the right position at the right time” is not a trivial process [48], since there is an imperative need for digital experts with different skills and background to fulfil a specific job role. Hence, the complex and multifaceted demand for digital workforce and sometimes, the blurry boundaries of job roles characterized by overlapping duties and responsibilities pose further recruitment challenges to O&G companies that are interested in acquiring professionals into their business processes but on the same time, they are not aware of what digital talent fulfils their expectations and needs. The objective of this RQ is the identification of prominent job roles across the examined technology trends and the definition of a formal portrayal regarding their responsibilities on the specific needs of the O&G industry.

- [RQ2.1] Which are the most prominent competencies and what are the most desired skillsets within each technology trend?
Motivation: As stated in RQ2, the digital transformation of the O&G industry is accompanied by continuous technological disruptions creating, in turn, a high demand for different skills.
In fact, a job opening defines a skillset of responsibilities and usually, interconnected competencies for an ideal candidate. In addition, these core hard skills refer to a broad taxonomy of developing tools, programming languages, platforms etc. synthesizing a profile of a future employee. RQ3 aims at the design of a competence map capturing the O&G industry interrelated requirements and trends in terms of the corresponding technological (or hard) skills.

IV. METHODOLOGY

In this section, we present, in detail, the methodology followed throughout the study in order to meet the objectives by addressing the posed RQs. The formulation of the research schema was dictated by previous research [59] that fully captures the scope of our goals and questions. The data-driven approach (Fig. 1) consists of seven phases that are (i) data collection, (ii) feature extraction, (iii) pre-processing, (iv) representation, (v) data analysis and (vi) presentation of the results and their implications to target groups and stakeholders.

A. DATA COLLECTION

Before proceeding to the detailed description of the proposed methodology, it is essential to define the core ingredient of the current study. To this regard, the basic unit of analysis is the job vacancy that can be thought as a semi-structured web document composed of a number of fields with a variety of entries. Bearing in mind the amount and complexity of information hidden in web job postings, we followed a similar approach presented in the study of Boselli et al. [51] in order to provide a more precise definition regarding the experimental unit of the study. Hence, a job vacancy $J$ can be perceived as a tuple of a finite number of ordered elements of the general form

$$J = (i, t, d, n, l, e, y, b, tr)$$

where $i$ is the identification number, $t$ is the title, $d$ is the description, $n$ is the company name, $l$ is the location, $e$ is the employment type, $y$ is the posting year of the vacancy found in job board $b$, while $tr$ is the trend to which the job vacancy belongs. At this point, we have to clarify that a job vacancy may belong to more than one technology trend.

A representative example of a job vacancy is presented in Figure 2, where there is plenty of information in both semi-structured and unstructured data. For example, the metadata fields company name, location and employment type provide certain details regarding the general characteristics of the offered job. In contrast, the title and more importantly, the description fields describe through free text the job role and the corresponding responsibilities and competencies for an ideal candidate to fulfill the job position. While not present in the example, the posting years could be found in the general description of a vacancy in job board $b$.

Focusing on the investigation of the demand for future workforce in O&G 4.0, we decided to collect vacancies posted at well-known dedicated job boards of this specific industry, so as to ensure a high coverage of relevant results. More precisely, two niche O&G job boards, namely the Rigzone\textsuperscript{5} and the Oil and Gas Job Search\textsuperscript{6} and a more generic one with specialized search option for O&G industry, the Energy Job Line\textsuperscript{7}, were selected as the main sources for retrieving data related to O&G job openings.

---

\textsuperscript{5} https://www.rigzone.com
\textsuperscript{6} https://www.oilandgasjobsearch.com/
\textsuperscript{7} https://www.energyjobline.com/
Title: Data Scientist
Company Name: XXXX
Location: Dhahran, Saudi Arabia
Employment Type: Full Time Salaried Experience

Description: We are seeking a Data Engineer to join Upstream Database Service Division (UDSD) under XXXX. UDSD provides and maintain upstream structured & unstructured data repositories, application middle-tier and Information Management services. Develop infrastructure & security applications for Upstream. Implement upstream data modeling, data security, data integration, data quality, data governance for enterprise data. Facilitate collaboration and provide consultation guidelines for business process management, and knowledge management services.

Minimum Requirements:
- The successful candidate should hold a bachelor's degree in Computer Science or equivalent.
- Minimum of 9 years' experience directly related to Data Modeling and Data Engineering.
- Expertise in database skills specifically SQL, complex querying, SQL explain plans and tuning.
- Excellent understanding of Data Security implementations in Oracle.
- Clear understanding of the concepts of Data Architecture supporting analytics.
- Expertise in data manipulation, data loading, data transformation, and data cleanup.
- Working knowledge of the PI Integrator, R language, and Oracle Advanced Analytics options.
- Knowledge of and experience with Hadoop implementations and NoSQL databases is desirable.
- Knowledge of Oil & Gas Upstream business desirable.
- Knowledge of and experience with Hadoop implementations, NoSQL databases is desirable.

Duties & Responsibilities: You are expected to be able to perform the following:
- Design and maintain the new Upstream Advanced Analytics database considering usage, security, performance, integrity, data migration, data transformation, and data archival.
- Build and maintain data models, which serve IR 4.0 projects.
- Implement data security, and set up the data streams, which feed the analytical database.
- Reflect changes in model from source and target data repositories, ensure data integrity, enforce data governance and guidelines for data analysis.
- Formulate policy to ensure proper data management and governance for Analytical Analysis database.

FIGURE 2. Illustrate example of job vacancy related to the digitalization of O&G industry

The collection phase was conducted through a python-based web crawler scraping all three job boards with the aim of retrieving relevant job vacancies associated to the examined technology trends. To this regard, the search strategy is a critical task that may considerably affect the quality and the inferential mechanism concerning the research goal of the current study. Hence, we followed a search plan involving the creation of three separate search strings consisted of specific keywords reflecting the three digital trends under investigation that are DA, IoT and Cloud. For the cases of IoT and Cloud technology trends, the definition of the appropriate keywords was rather a simple process, since we opted to utilize “umbrella” terms such as “internet of things”, “industrial internet of things” (along with their corresponding abbreviations “iot” and “iiot”, respectively) and “cloud”. On the other hand, DA is certainly a wide-ranging concept that comprises a broad terminology related to data and the corresponding data life cycle phases. For this reason, the keywords for DA were defined after an iterated approach based on trial searches through representative terms (i.e. “data analysis”, “data management”, “data integration”, “data visualization” etc.) based on the empirical evidence from other similar studies. Finally, the collection process of job vacancies was conducted at two distinct timestamps covering the years 2019 and 2020 after applying the corresponding search string describing each technology trend.

| Name   | Description                                                                 |
|--------|-----------------------------------------------------------------------------|
| id     | ID of the job                                                               |
| title  | Title of job                                                                |
| description | Description of the job                                 |
| company | Company posting the job                                                     |
| location | Location of the job                                                       |
| region | The geographical region of the job (Asia-Pacific, Europe, Latin America, Middle East, North America) |
| employment type | The employment type of the job                                           |
| year   | The posting year of the job                                                |
| board  | The board used for the identification of job (Rigzone, Oil and Gas Job Search, Energy Job Line) |
| trend  | The technology trend according to search string (DA, Cloud, IoT)           |

B. FEATURE EXTRACTION
The data fusion process resulted into a unified database containing a collection of job vacancies from the three dedicated job boards. Due to the fact that the information in each job vacancy is provided through both free text (title and description) and several metadata fields (company name, location and employment type etc.), this step involves their transformation into a more structured format. Initially, the web
crawler utilized for the retrieval of the data browsed through HTML tags the content of job boards and stored necessary information in lists, with each list representing a job vacancy and its fields. Lists were then transformed to separate data frames and concatenated to produce the final unified database of job postings that was used in later stages as a starting point of analysis. In Table 1, we present the features extracted and stored during the crawling process.

C. PRE-PROCESSING

After the feature extraction phase, important cleaning and pre-processing steps were undertaken to remove noise and irrelevant information. Initially, non-English job postings were removed, as they were deemed to be targeted to a narrower audience, thus limiting the scope of our analysis. The final dataset, after removing identical postings found in more than one job board (deduplication process) is comprised of 1999 vacancies.

Regarding the textual content of the extracted job vacancies (i.e. title and description), we applied appropriate cleansing and pre-processing techniques. Initially, textual data were reverted to lowercase, while also removing punctuation marks, special characters, URLs and delimiters. In addition, stop-words and whitespaces were also removed, while each posting was tokenized, stemmed and lemmatized. During this phase of the methodology, the NLTK python library was used.

D. REPRESENTATION

The representation of unstructured textual content found in both the title and description fields and the extraction of meaningful information are certainly not so trivial tasks, since they involve the adoption of Natural Language Processing (NLP) and, by extension, Text Mining (TM) methodologies. The pre-processing procedure, described in the previous section, constitutes in fact, a prerequisite step for the representation of both textual features into appropriate NLP models.

1) JOB TITLE FEATURE

As we have already mentioned, the title of each vacancy summarizes in a compact way important information related to a job opening and more specifically, it reflects the level and the general responsibilities of the position. On the other hand, the relative short textual content, and sometimes the informal way of the writing style without providing a clear progression of the job role, along with the compressed information found in title field pose significant challenges to the extraction of useful outcomes.

To meet the objective of RQ2, each title was transformed into a set of n-grams, where a n-gram can be defined as a contiguous sequence of n terms from a given sequence of text [60]. In particular, we decided to evaluate a set of n-grams with a varying number of terms in order to identify meaningful job roles. Finally, after an evaluation process for the set of identified n-grams and their number of occurrences, we concluded that bi-grams (2-grams), defined as a sequence of two terms, provided a list of essential job roles for each technology trend. In order to grasp the notion of a bigram, the title of the demonstrative example (Figure 2) results into the bigram “[Data, Scientist]”.

2) JOB DESCRIPTION FEATURE

In contrast to the title feature, the description field constitutes a wealth source of hidden information related to specific aspects of a job opening such as responsibilities, and on some occasions, list of compensations and benefits for a job seeker. More importantly, this specific field describes the desired competencies and expertise that candidates need to carry out a specific job. In turn, the identification of skillsets through the analysis of specific labour market needs provides a straightforward mechanism for the effective competence mapping across different industrial sectors and job roles.

As we have already mentioned, one of the primary objectives of the current study is the identification of the set of hard skills across different technology trends that are demanded for the transition into the new digitalized O&G era. Although the building of competence maps could bring several benefits to a wide range of target groups (O&G industry leaders, governmental and educational institutions, students, job seekers, current employees etc.), the definition of the required and continuously varied skills in such a dynamic labour market environment is not a trivial task. In addition, despite the fact that job vacancies provide a basis for the construction of competence maps, the unstructured nature of data along with the plethora of technological skills are significant barriers to the extraction of job profiles.

To overcome the above limitations and challenges, we propose an approach that identifies hard skills through the appropriate matching of terms found in documents highlighting the desired skillsets according to a predetermined lexicon (or reference list) [59]. This step, in turn, results in the representation of job description’s textual content into a multi-dimensional Vector Space Model (VSM) with Boolean terms. More precisely, the basis for hard skills extraction is the construction of a lexicon comprising a list of technologies formulated by the Developer Surveys launched from Stack Overflow and taking into consideration the period 2014-2020. The concept behind exploiting the Stack Overflow Developer Survey in contrast to other technological taxonomies is attributed to the fact that Stack Overflow is one of the most respected and well-known ICT communities dedicated to knowledge-sharing [61]. Additionally, as the community’s main function is to allow developers to post their questions and answers regarding ICT topics, it is a direct and rich source of the hard skills exploited by professionals. Apart from the purposes of the platform, the Stack Overflow Developer Survey ensures extensive coverage of topics related to a wide

---

8 https://www.nltk.org/  
9 https://insights.stackoverflow.com/survey/2020
range of ICT, with 65,000 developers and experts providing their preferences based on their practical experiences. Moreover, the survey leverages an analytical separation of technologies by not only mentioning hard skills but also distinguishing them into more generic categories regarding their purpose of use. Thus, the Developer Survey was deemed the dominant source for the identification process of hard skills found in job postings.

Described briefly, the lexicon can be conceptualized as a reference competence taxonomy composed of two tiers. The First Tier encompasses seven Competency Classes (CC) defining more general knowledge and expertise related to different aspects of ICT, whereas the Second Tier is comprised of explicit 182 Hard Skills (HS) referring to specific ICT technologies (Figure 3). Regarding the CCs of the First Tier, the category of Languages\textsuperscript{10} refers to programming tools used to develop code (e.g., Python, JavaScript, R). Web Frameworks refer to specialized, self-contained environments devoted to the construction and deployment of front-end and back-end architectures (e.g., Angular). The Big Data category contains technologies that are employed to process large amounts of data and develop machine learning models for specific purposes. Developer Tools and Collaboration Tools concern dedicated Integrated Development Environments (IDEs) widely used for developing code and software sharing suites, respectively. Operating systems, virtual environments and hosting services are included in the Platforms category, while database management systems are contained in the Databases category.

![FIGURE 3. Competence taxonomy for the identification of hard skills](image)

The next step involves the matching of the lexicon terms found in the description fields, with the intention of transforming the textual content into a vector space representation [59]. The Vector Space Model (VSM) for a document $D$ and a set of terms $V = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n\}$ can be defined as a high dimensional vector $x \in R^n$, associated with the document, with the length of the vector being equal to the length of $V$. Each element $x_i$ corresponds to a statistical metric related to the occurrence of each term of $V$ inside document $D$ [62]. Thus, the produced vector reflects the matching of terms from $V$ inside each document $D$. A common representation for vector elements is the binary set $\{0,1\}$ dictating the presence (denoted by 1) or absence (denoted by 0) of a term in a specific document $D$. In the case of our study, each vacancy is represented by a high dimensional vector with 182 elements corresponding to HS that are present or absent in the description field. Thus, the VSM representation is accomplished through a matching process of HS terms detected in each vacancy description to the predefined terms of the reference lexicon.

An example of the VSM representation process can be demonstrated on the description of the job vacancy of Figure 2. Based on the CC of the First Tier defined in Figure 3 and the proposed matching process, the identified HS distributed across their respective classes are SQL (Databases), Oracle (Databases), R (Languages), Hadoop (Big Data) and NoSQL (Databases). Thus, the produced vector of this vacancy presents ones (1s) for all elements corresponding to these technologies and zero values (0s) to all remaining terms.

The information extraction process for the *description* field was completed by assembling all vectors into a term matrix, where each row represents a job vacancy and its corresponding VSM representation and each column a HS term from the competence taxonomy of Figure 3 (Second Tier).

**E. DATA ANALYSIS**

After the representation of extracted features into a readable format, we focus on the application of appropriate data analysis methodologies for providing answers to the RQs (Section 3) and derive meaningful conclusions. Table 2 provides a detailed mapping between RQs and features extracted from the collection of job vacancies accompanied by the data analysis and visualization techniques used in this study.

Generally, for RQ$_2$, we made use of the metadata features and appropriate univariate statistical methods (descriptive statistics and graphical representation of distributions) for summarizing meaningful conclusions regarding the characteristics of job vacancies. In addition, to gain better insights concerning the current labor market demand and uncover interesting patterns and trends, we utilized statistical inferential procedures. To test the association between the characteristics of job offers, the chi-square test of independence was performed. Regarding RQ$_{1.2}$ that involves the exploration of the geographical distribution of job postings, we focused on *location* metadata feature. Even though it is not an obligatory field, as some companies support

---

\textsuperscript{10} For the rest of the paper, the competency classes of the First Tier are highlighted by bold fonts.
remote working or require their staff to frequently travel to other regions, most organizations chose to include the location as supplementary information. The extracted locations were transformed from simple text to latitude and longitude coordinates via the use of a Geocoding platform\textsuperscript{11}. Geocoding results were then mapped by the Leaflet\textsuperscript{12} library in interactive heatmaps. Finally, Multiple Correspondence Analysis (MCA) \textsuperscript{[63]} was performed on a specific set of extracted metadata features (technology trend, year, location) related to RQ\textsubscript{1} to get a comprehensive overview regarding the characteristics of the current digital demand.

\begin{table}[h]
\centering
\caption{Data Analysis Methods for RQs and Extracted Features}
\begin{tabular}{|l|l|l|}
\hline
Research Question & Feature used & Analysis Method \\
\hline
RQ\textsubscript{1,a} Which are the characteristics of job offers in terms of technology trends, employment type and year of job opening? & technology trend, employment type, year & Descriptive statistics, inferential statistics (chi-square test) \\
RQ\textsubscript{1,b} How are these job offers distributed at geographical level across the technology trends? & location & Heatmaps, descriptive statistics, (chi-square test) \\
RQ\textsubscript{1,c} Who are the leading players (companies) driving the digital transition into the O&G 4.0 era? & company & Descriptive statistics, inferential statistics \\
RQ\textsubscript{1,d} Which are the digital job roles with the highest demand in the O&G 4.0 era? & title & Natural language processing, text mining, bigrams \\
RQ\textsubscript{1,e} Which are the most prominent competencies and what are the most desired skillsets within each technology trend? & description & Co-term analysis, graph theory, community detection (Louvain algorithm) \\
\hline
\end{tabular}
\end{table}

The objective of RQ\textsubscript{2} involves the distinction of ICT professionals into separate job roles, with each role being leveraged into a specific application domain of the O&G industry. To that end, we took into consideration the categorization of O&G services to upstream, midstream and downstream with the aim of defining profiles and respective duties to the extracted job roles from the title field. With this profiling process, our main scope was to capture the essence of each role and examine its connectivity to one of the three technology trends. The extracted bigrams (see Section 4.4.1) were the core elements of the analysis as the evaluation bigram’s frequencies contribute to the identification of highly desired positions. Definitions for all positions were assigned by a review-based process, where each job underwent a thorough examination to discover its functionality and contribution to the multifaceted services of the O&G industry.

Concerning RQ\textsubscript{s}, our purpose was first, to extract the most prominent HS required for fulfilling a job vacancy and then, to design competence maps that summarize desired HS along with their interconnections for each one of the three technology trends. To this regard, we investigated the co-occurrences of HS and then, we made use of Graph Theory methods to facilitate the inferential process \textsuperscript{[64]}. The rationale behind our approach is that HS (or terms) appearing together (or co-occurring) in the job description field can be used for the mapping of prominent competencies belonging to a specific technology trend. To this regard, the evaluation of appropriate indices based on co-occurrences of terms are used to measure the strength of associations between detected HS. Finally, the results of the co-term analysis constitute the basis for the identification of clusters encompassing a subset of interconnected and emerging HS required for each technology trend.

Based on the above considerations, a critical decision concerns the choice of an appropriate similarity measure able to capture in a satisfactory way the degree of association between pairs of HS terms. During the past decades, there has been proposed a variety of both simple and more sophisticated indices for analyzing the co-occurrences of terms extracted from the corpus of textual content \textsuperscript{[65]}. Callon et al. \textsuperscript{[66]} point out that a simple counting of co-occurrences is a naïve approach that can lead to biased findings and for this reason, they strongly suggest the utilization of probabilistic similarity measures. In the current study, we made use of the equivalence index (EI) \textsuperscript{[67]}, since it can be considered as the most appropriate measure taking into account the probabilistic nature of co-occurrences of terms \textsuperscript{[65]}. Adapting the definition of EI to the needs of the current study, it can be evaluated through the following formula

\[
E_{ij} = \frac{c_{ij}^2}{c_i c_j}
\]

where \(c_{ij}\) is the number of job vacancies in which two hard skills \(i\) and \(j\) co-occur and \(c_i\) and \(c_j\) represent the occurrence frequency of hard skills \(i\) and \(j\) in the set of job vacancies, respectively. EI ranges into the interval \([0,1]\) with zero value indicating the perfect dissimilarity of two hard skills \(i\) and \(j\), or in other words, these specific hard skills are never required both as desired competencies for fulfilling a job opening. In contrast, an EI value of one indicates that two specific hard skills \(i\) and \(j\) are always appear together in the description of a job opening and thus, they can be considered as a perfect pair of hard skills for a specific job.

Despite the efficiency of EI in highlighting the strength of association between desired HS, the evaluation of all possible pairs results in a high amount of information that may be impractical for inferential purposes regarding the design of competence maps. For example, a low value of EI implies pairs of HS that are seldom appear together and thus, they cannot be considered as good candidates for the design of a competence map for a specific technology trend. Certainly, this type of information might provide meaningful

\textsuperscript{11} https://developer.here.com/projects/

\textsuperscript{12} https://leafletjs.com/
conclusions regarding for example the detection of rising HS, but this is out of the objectives of the current paper. Thus, we focused our interest on the detection of a subset of HS presenting a relatively high degree of association and for this reason, they can be considered as skillsets within a technology trend. To this regard, we made use of advanced methodologies from Graph Theory and community detection through community detection algorithms. Generally speaking, a competence map graph is in fact, a network, where nodes represent HS that are linked together through edges produced by the EI values for each pair of HS. The next challenge involves the identification of communities (or sub-graphs) within networks that contain HS that are frequently appeared together and can be considered as skillsets for a technology trend.

For this aim, we apply a well-known community detection method, namely, the Louvain algorithm [68] which relies on network modularity to detect communities. In brief, modularity ($Q$) is a metric that measures the quality of the network’s division to communities and more precisely, the relative density of edges inside communities with respect to edges outside communities [69]. The formulae for the evaluation of modularity is given below

$$Q = \sum (e_{ii} - a_{ii}^2)$$

where $e_{ii}$ is the number of edges belonging in a community $i$ and $a_{ii}^2$ is the number of edges that connect nodes from community $i$ of the network with another nodes of other communities in the network. Modularity ranges into the interval $[-0.5, 1]$ and provides a straightforward way to evaluate the tendency of a network to either be comprised of disconnected nodes, in the case of negative values, or to form connected components, in the case of positive values. The Louvain method is, in fact, an optimization algorithm based on the maximization of the objective function, called network modularity, by an iterated approach through the assignment of new nodes in communities until the modularity of a network reaches its highest possible value [68].

### RESULTS

In this section, we present the findings of this study based on the posed RQs.

**[RQ.1]** What is the current demand for digital workforce in the O&G 4.0 era?

**[RQ.1.1]** Which are the characteristics of job offers in terms of technology trends, employment type and year of job opening?

Table 3 summarizes the characteristics of job offers based on information extracted from metadata. The exploration of the distribution regarding the examined technology trends indicates that there is a noteworthy demand for talented workers empowered by DA skills (52.5%), whereas the percentages are lower for technologies such as Cloud computing (34.9%) and IoT (12.6%). One possible reason for this observed divergence in demand may be due to the fact that DA, compared to Cloud and IoT, is a more generic technological concept that encompasses a broad and sometimes even hard to distinguish branch of job roles, profiles and corresponding competencies.

| Variable          | Levels         | $N$ | %  |
|-------------------|----------------|-----|----|
| Technology trend   |                |     |    |
| DA                | 1049           | 52.5|    |
| Cloud             | 698            | 34.9|    |
| IoT               | 252            | 12.6|    |
| Total             | 1999           | 100.0|   |
| Employment type   |                |     |    |
| Permanent         | 1544           | 92.3|    |
| Contract          | 124            | 7.4 |    |
| Contract or Permanent | 4          | 0.2 |    |
| Part Time         | 1              | 0.1 |    |
| Total             | 1673           | 100.0|   |
| Year              |                |     |    |
| 2019              | 1156           | 57.8|    |
| 2020              | 843            | 42.2|    |
| Total             | 1999           | 100.0|   |
| Region            |                |     |    |
| North America     | 725            | 38.5|    |
| Europe            | 535            | 28.4|    |
| Asia-Pacific      | 370            | 19.7|    |
| Middle East       | 151            | 8.0 |    |
| Latin America     | 97             | 5.2 |    |
| Africa            | 4              | 0.2 |    |
| Total             | 1882           | 100.0|   |

Regarding the employment type, more than 92% of job openings provide a full-time permanent position indicating a shift in the O&G industry for technology adoption through the attraction, recruitment and team building of talented digital workforce. Finally, there is a notable difference in job offers between the two examined time snapshots (2019 and 2020), which might be caused by the detrimental effects of the Covid-19 crisis and the recruitment shortcuts on the labor force worldwide.

**[RQ.2]** How are these job offers distributed at geographical level across the technology trends?

Concerning the geographical distribution of job offers, Figure 4 illustrates the worldwide labor market needs across the three technology trends. The variation in color (from light blue to red) of the circles provides a straightforward visual manner to explore how job openings vary over regions. We have to note that the geolocation entries are extracted at the zip code level after omitting job vacancies without providing such information. At a higher level of aggregation (Table 3), it is evident that North America region provides great career opportunities (almost 39% of total vacancies) for job seekers that are willing to take an active role in the digitalization of the O&G industry.

The finding for this specific region is rather than rational, if we take into consideration the plethora of cutting-edge technological companies offering pioneering products, services, and solutions across key industries. Moreover, companies located in Europe offer also a relative high percent of O&G job opportunities (28.4%) and thus, they seem to be confronted to the current needs of digital transition followed by Asia-Pacific, a region with a rapid economic and employment generation growth. Finally, there is also a remarkable proportion of job vacancies in Middle East region,
a fact that is certainly associated to the quantity and quality of O&G resources and the world leaders settled and operating in this specific region.

![Data Analytics](image1)

![Cloud](image2)

![IoT](image3)

**FIGURE 4.** Geographical distribution of job vacancies across technology trends

[RQ1.3] **Who are the leading players (companies) driving the digital transition into the O&G 4.0 era?**

Despite the fact that the digitalization of the O&G industry is characterized as a generally slow and low-maturity process due to the technological and organizational limiting factors (see Section 1), leading companies seem to realize the potential value of this digital transition. To this regard, both multi-national energy giants (e.g. Baker Hughes, Shell, ExxonMobil etc.) and professional services companies (e.g. PwC) has shifted to a new way of digital thinking by adopting technological initiatives in their operational processes (Figure 5). In order to accomplish this challenging task, they have adopted a transition model by investing in human capital, a fact that is indicated by the rising of job openings related to several operational processes of the O&G industry.

Indeed, the empirical evidence reveals that several companies have engaged in initiatives aiming to invest in the digitalization of their services and the overall training of their workforce in innovative new technologies. Baker Hughes launched the Digital Solutions platform\(^ {13}\), offering remote inspection and monitoring in the O&G services along with performance evaluation and product assessment, in an effort to improve its functionalities. Shell is also heavily supporting digitalization, investing in Blockchain technology and Artificial Intelligence\(^ {14}\) and recently partnering with Microsoft to encourage green energy practices combined with technological prowess\(^ {15}\).

Complementary actions are also being taken in order to ensure proper digitized recruitment in the company\(^ {16}\). Similar partnerships have been formed with ExxonMobil\(^ {16}\) with increased focus on leveraging IoT and data-driven frameworks. ExxonMobil has also proceeded to enterprise-based software in order to boost its productivity and revenue\(^ {17}\). PwC, in an effort to design strategies for bridging the gap between digital services and physical operations in O&G, has accumulated data from 200 O&G companies to review their digital activity\(^ {18}\).

\(^{13}\) https://www.bakerhughesds.com/

\(^ {14}\) https://www.shell.com/energy-and-innovation/digitalisation.html

\(^ {15}\) https://www.shell.com/energy-and-innovation/digitalisation/news-room/shell-and-microsoft-form-alliance-to-help-address-carbon-emissions.html

\(^ {16}\) https://energyfactor.exxonmobil.com/news/a-data-transformation-in-americas-energy-frontier/

\(^ {17}\) https://www.appsruntheworld.com/customers-database/customers/view/exxonmobil-united-states

\(^ {18}\) https://www.strategyand.pwc.com/gx/en/insights/2020/digital-operations-oil-and-gas.html

**FIGURE 5.** Distribution of top 10 companies offering digital job
Based on the results, only 7% of the studied companies report digital maturity and increased technological support. In the same report, the main technologies that can help shaping a new digital landscape for O&G companies are Cloud Computing, IoT and Machine Learning, indicating the need for specialised workforce that can respond to this pressing need. BP, in collaboration with Accenture\(^1\), ventures in Artificial Intelligence initiatives \(^2\) and robotics in order to monitor O&G production, having impressive results on its efficiency and annual revenue \(71\).

After the examination of the labor market demand for digital workforce, the next step involves the application of inferential mechanisms with the aim of detecting interesting patterns related to these characteristics. For this reason, we graphically explored the joint distributions of the extracted metadata fields, whereas the chi-square test of independence \(\chi^2\) was omitted from any further analysis, since it presents a negligible number of job offers, reveals a statistically significant association, \(\chi^2(8) = 78.938, p < 0.001\). In particular, the job openings seem to be similarly distributed across the examined regions (Figure 6b), except Middle East, where the demand for DA related jobs (85.4%) dominates. Finally, the chi-square test indicated a statistically significant association between the number of identified job offers at the two examined timestamps and the region, \(\chi^2(4) = 131.093, p < 0.001\). The investigation of their joint distributions (Figure 6c) highlights a significant increase of digital demand in Asia-Pacific and Latin America labor market compared to the previous year.

Figure 7 presents the graphical representation of the results derived from MCA. Regarding the interpretation of the MCA plot and the relative positions of points, we have to note that levels of different variables that are ordinated close to one another designate similar job characteristics. In addition, profiles of job vacancies that are very different from the average profile, are located far from the origin. The graphical inspection reveals that the IoT technology trend, the two levels of Year and the Asia-Pacific and Latin America regions contribute the most to the definition of the first dimension. In contrast, the Middle East region and the two remaining technology trends (DA and Cloud) mainly contribute to the definition of the second dimension. Based on the interpretation guidelines, an interesting finding concerns job vacancy gathered at 2020 demonstrating the increase on digital workforce related to IoT technology trend, since the latter is represented by an extreme point far from the origin and the corresponding level of 2019 (opposite direction of \(x - axis\)).

Furthermore, the proximity of the points Asia Pacific and Latin America regions to 2020-year point also demonstrates the increase on these specific labour market regions, despite the overall decreasing trend worldwide. Finally, the MCA plot clearly supports the evidence of the DA dedicated labour market at the Middle East region.

\(1\)https://www.accenture.com/us-en/case-studies/energy/empowering-productivity-digital-transformation

\(2\)https://www.bp.com/en/global/corporate/what-we-do/technology-at-bp/digital-innovation.html
[RQ2] Which are the digital job roles the with highest demand in the O&G 4.0 era?

Given the fact that the digitalization of the O&G industry refers to a broad spectrum of technological innovations covering different aspects of operational and business practices and processes, in RQ2, we focus on the identification of essential job roles within the O&G 4.0 era. As we have already described, we made use of NLP and TM methodologies (Section 4.5) on textual content found in the title field of the collected job vacancies. Figure 8 summarizes the top five job roles across the three technology trends as identified through the extraction of bigrams based on textual content of job titles. Although there is a proportion of job roles that are common for all technology trends, there is also a subgroup of job roles aligned to the special needs and responsibilities of each digital trend.

Generally speaking, DA trend encompasses roles that are mostly related to the interdisciplinary domain of Data Science and its wide range of corresponding responsibilities. A noteworthy finding is the increase demand for Software Engineers to fulfill job roles related to DA technology trend. Despite the fact that both roles involve advanced programming skills, there is also a fundamental distinction between these two roles, since Software Engineers are responsible for the design, development and maintenance of software rather than data-oriented processes. Regarding the remaining technology trends, the identification of top job roles indicates that companies interested in the adoption of Cloud and IoT products and services seek for IT professionals with a strong focus on Software Engineering principles.

To clarify the duties of each role and further highlight the differences among them, we provide a brief description of the responsibilities of each position, tailored to the upstream (extraction and drilling), midstream (transportation and storage) and downstream (transformation to useful products) services of the O&G industry. These characterizations originate from the descriptions of vacancies that contained such roles and capture the key duties of a specialist belonging to them.

Table 4 summarizes the most prominent HS across the three technology trends, according to their frequency of appearance and the competence taxonomy presented in Figure 3. A percent cut-off value of 5% was dictated in order to present only the most desired expertise. As made evident, a competency profile for professionals belonging to each trend can be extracted, with skills matched to the CC that they are associated to.

The DA trend is heavily dependent on HS related to Languages (First Tier of taxonomy) which reveals the need of job roles in this area (e.g. Data Scientists, Data Engineers, Data Analysts etc.) for producing and maintaining code for various purposes. The Cloud trend appears to be significantly more diverse, containing HS from almost any CC, something possibly attributed to the multidisciplinary nature of cloud solutions which combine software development (Web Frameworks) with security and database systems (Databases) and code sharing through Collaboration Tools. Finally, the IoT trend leans towards increased language-related competencies (Languages) but appears to demand platform-based competencies (Platforms) as well. This finding can be justified, if we take into consideration that IoT depends on inter-platform communication and data transferring protocols.

Generally speaking, the demand for the examined technology trends is characterized by common HS, which are marked with grey colour on the table. However, depending on the trend, and by extension to the job roles associated with it, the gravity of HS diverges, as their knowledge by professionals is either more or less desired. For example, while the R programming language holds a high position in the DA trend, being a high-end tool for data scientists, it presents significantly lower ranking for the remaining two trends. Similar conclusions are extracted from the analysis regarding the Java language, which is held in high demand in IoT and particularly in Cloud job positions but on the same time, it is not heavily demanded in DA.
FIGURE 8. Description of job roles per trend

| TABLE IV | MOST POPULAR COMPETENCIES (HARD SKILLS) FOR EACH TECHNOLOGY TREND |
|----------|---------------------------------------------------------------|
| **DA**   | **Cloud**                                                     | **IoT**                                       |
| HS (2nd Tier) | CC (1st Tier) | N (%) | HS (2nd Tier) | CC (1st Tier) | N (%) | HS (2nd Tier) | CC (1st Tier) | N (%) |
| SQL DB   | 26.79         | AWS P  | 26.36         | SQL DB       | 21.83   |        |        |        |
| Python L | 25.17         | Python L | 24.36        | C L          | 21.03   |        |        |        |
| C L      | 17.16         | SQL DB | 23.78         | Python L     | 20.63   |        |        |        |
| R L      | 15.44         | Java L  | 22.92         | Java L       | 16.67   |        |        |        |
| Java L   | 14.01         | C L     | 18.62         | Oracle DB    | 11.90   |        |        |        |
| Oracle DB | 12.77         | JavaScript L | 13.18       | R L          | 10.32   |        |        |        |
| AWS P    | 9.15          | JavaScript L | 14.76       | Git CT       | 9.92    |        |        |        |
| JavaScript L | 7.34         | Git CT  | 13.18         | AWS P        | 8.73    |        |        |        |
| C++ L    | 6.67          | Linux P | 11.89         | JavaScript L | 8.33    |        |        |        |
| Go L     | 6.10          | Windows P | 10.89        | Linux P      | 8.33    |        |        |        |
| Hadoop BD | 6.01          | Go L    | 10.32         | .NET DT      | 7.94    |        |        |        |
| Windows P | 5.72          | R L     | 9.03          | C++ L        | 7.54    |        |        |        |
| Linux P  | 5.62          | Docker P | 7.74          | Windows P    | 7.14    |        |        |        |
| Git CT   | 5.15          | Microsoft Azure P | 7.16       | Hadoop BD    | 5.56    |        |        |        |
| PowerShell L | 7.02         | PowerShell L | 5.16        |        |        |        |        |        |
| C++ L    | 6.73          | Docker P | 5.16          |        |        |        |        |        |
| Angular WF | 6.59          | Kubernetes P | 5.16        |        |        |        |        |        |
| Ansible CT | 6.59          |        |        |        |        |        |        |        |
| HTML L   | 6.16          |        |        |        |        |        |        |        |
| CSS L    | 6.02          |        |        |        |        |        |        |        |
| Kubernetes P | 6.02          |        |        |        |        |        |        |        |
| .NET DT  | 5.73          |        |        |        |        |        |        |        |
| Salesforce P | 5.59          |        |        |        |        |        |        |        |
| Node.js WF | 5.16          |        |        |        |        |        |        |        |

**Note**: L = Languages, WF = Web Frameworks, BD = Big Data/ML, DT = Developer Tools, CT = Collaboration Tools, P = Platforms, DB = Databases
Some skills are actively sought by all three trends, such as SQL, but in general, each trend requires increased expertise in specific competencies in order to support the job roles adhered to it. In addition, apart from common HS, some trends require proficiency in certain skills such as Docker and Kubernetes (Platforms) and Web Frameworks related competencies for the Cloud trend (Angular, Node.js) or Hadoop for the DA trend. This is an anticipated outcome, as job roles tend to be divergent from each other and have different duties, despite possible similarities between them. Overall, the Cloud trend appears to demand the most exclusive expertise, potentially explained by its multitargeted aspects, which take advantage of several technological advances.

After the identification of prominent HS for every technology trend, we concentrated on the exploration of their interconnections within each technology trend. Our scope is to gain deep insights regarding the demand for specialized groups of hard skills (or skillsets) that a candidate should have in order to fulfil a job opening in a specific technology trend. To this regard, we graphically explore the interconnections of HS through the utilization of graph theory and network construction. In practice, HS are represented by network nodes, whereas edges represent the degree of association between pairs of nodes expressed by their equivalence index (Section 4.5). Finally, the application of the community detection algorithm resulted to the identification of groups of skillsets within each technology trend. Hence, in our study, a skillset $i$ ($S_i$) can be defined as a set of HS identified by the community detection algorithm representing a group of associated competencies required for a specific technology trend.

**Skillsets for DA:** A demonstrative example of the detected skillsets for DA technology trend is presented in Figure 9, in which Louvain algorithm identified 16 communities (or skillsets). Each skillset comprises a group of interconnected HS from the Second Tier of taxonomy (Figure 3) that may belong to the same or different CC of the First Tier. Having a closer inspection on the structure of the competence network and its communities, we can extract a few noteworthy findings regarding this specific technology trend.

More specifically, Skillset 1 (denoted as $S_1$ in Table 5) combines HS from various CC (Languages, Web Frameworks, Big Data, and Databases) related to all phases of data science lifecycle for example extraction from web content (e.g., jQuery, Spring), storage (i.e., Oracle) and retrieval of data (e.g., SQL) from databases, data analysis (e.g., R, Python, Matlab) and visualization of data on web applications (e.g., CSS, HTML). Having in mind the mass amount of accumulated data and the need for specialized technologies for processing too large and complex datasets, this skillset also encompasses expertise in big data technologies (i.e., Hadoop). Generally, this competence profile is related to general duties and responsibilities of Data Scientist and Data Analyst job roles (Figure 9).

**Skillsets for Cloud:** The skillsets extracted from the community detection for the two remaining technology trends are summarized in Table 5. Regarding the Cloud technology trend, a first remarkable conclusion is the fact that it is notably divergent from its DA counterpart, as detected skillsets appear...
to be more multidisciplinary, containing HS from several CC. This outcome is somewhat expected as Cloud job vacancies frequently require expertise from various domains of ICT, such as Languages, Databases or Platforms.

To this regard, Skillset 1 contains Languages (Typescript, Java, SQL, C, C++, Python, Jquery etc) accompanied not only by Web Frameworks (Node.js, Angular, Asp.Net) but also by Platforms and Databases for hosting and deployment (AWS, Oracle). Moreover, Skillset 1 includes the most prominent HS of this technology trend (Table 4) that is Amazon Web Services (AWS), a fact that signifies the high demand for the development of cloud technologies in combination with Languages specialized in data analytics (Python) and data manipulation and storage (SQL). Based on the competence profile comprised by Skillset 1, a Full Stack Developer would be the most likely job role profile with these HS.

Certainly, the Cloud trend is more dedicated on technological advances that are related to the development of web applications and hosting services but it is not limited on this purpose, since there is an imperative need for expertise in data handling and security protocols development. Tools used for such purposes are contained in Skillset 2. This skillset is comprised of data-oriented Languages (R, Scala, Kotlin, Matlab) and Big Data tools (Hadoop, Keras, Tensorflow, Pytorch). In addition, these data-driven solutions frequently require servers for hosting and rely on the development of specialised software for data services and security protocols to ensure data privacy. Thus, in this skillset, there are also technologies of the general classes Platforms such as Docker and Kubernetes, completing the competence profile of a CyberSecurity Professional or a Software Engineer.

Similarly to DA, Skillset 3 contains primarily, Databases (Cassandra, CosmosDb, MongoDB etc), Big Data frameworks (Apache spark) and specific cloud Platforms (Couchbase, Amazon Echo) emphasizing the need for Data Engineers job roles with a strong focus on organising and monitoring the efficient creation of data storages. Skillsets 4 and 5 appear to be more web-oriented and correspond to the demand for Software Developers. Web Frameworks (Drupal, Flask, Dgango) and Languages widely used in Software Development (PHP, Perl, Ruby) along with code-sharing Collaboration Tools (Github, Gitlab, Ansible) are constituting the core competencies of these skillsets corresponding to the need of the Cloud technologies to develop software suitable for data manipulation and support.

**Skillsets for IoT:** As far as IoT technology trend concerns, it is substantially different from the previous ones, as skillsets are fewer and appear to be more intertwined. Having in mind that IoT is still in a transition phase and few attempts have been taken place in defining the key technologies that characterise this trend [72-74], it is expected the derived skillsets will combine various CC.

Skillset 1 reinforces this general belief, since we can observe that it contains HS related to data oriented Languages (e.g. Scala, Python, SQL etc.), Databases (i.e. PostgreSQL, MySQL) and Platforms (Kubernetes) in combination with Big Data expertise (i.e. TensorFlow, Keras etc.). Skillset 2 is similar to Skillset 1, containing Languages such as R, but also delves deeper to operating systems management with Platforms (Windows, Linux) and technical Languages also delves deeper to operating systems management with Platforms (Windows, Linux) and technical Languages (Bash). While Skillset 1 would characterise a Data Engineer, Skillset 2 could also be used to describe the competence profile of a Software Engineer. As IoT has been widely associated to mobile development, Skillset 3 contains mobile Platforms (Android, IOS) along with Languages such as Typescript and C# that can support the development of application for this generic purpose. Skillset 4 concerns Web Development, as it contains relevant Languages (HTML, CSS), Web Frameworks (Spring, Angular) along with hosting Platforms (Microsoft Azure,Predix) and can possibly be attributed to the development of tools supporting the IoT infrastructure. Both of these skillsets refer to candidates who wish to evolve as Software Developers and Design Engineers. In addition, Skillset 5 is unique in its nature, heavily focusing on Collaboration Tools and monitoring Platforms, effectively corresponding to the competence profile of a Program Manager.

An interesting finding is that Skillset 9 contains Assembly, a specific Language primarily used for hardware development. Considering that sensors and hardware toolkits such as Arduino are vital for IoT development, this finding confirms the multifaceted concept of IoT, which encapsulated data processing tasks with both software and hardware development.

**VII. THREATS TO VALIDITY**

In this section, we present and discuss potential threats to the validity of our study focusing on construct, internal/external validity and reliability. Regarding construct validity, described as the degree in which the examined phenomena are observed and measured, the data collection was based on a semi-automated strategy. Firstly, the search terms related to the examined digital trends were carefully calibrated to meet the objectives of the study. The final search strings for the cases of IoT and Cloud trends were quite generic to retrieve as many relevant jobs as possible eliminating the bias of the search process. This is attributed to the fact that job vacancies belonging in these trends can be adequately expressed with these search strings, as their titles and responsibilities are more specific.
| TABLE V | DETECTED SKILLSETS (COMMUNITIES) FOR EACH TECHNOLOGY TRENDS |
|---------|-------------------------------------------------------------|
| **Skillsets (DA)** | **Skillsets (Cloud)** | **Skillsets (IoT)** |
| $S_1 = \{SQL, R, Python, C, Java, JavaScript, Matlab, PHP, CSS, HTML\}_{L}$, $\{jQuery, Spring\}_{WF}$ | $S_1 = \{CSS, HTML, JavaScript, Java, C, SQL, Python, TypeScript, Vue, .NET, ASP.NET, Angular\}_{WF}$, $\{AWS\}_{P}$, $\{Git\}_{CT}$, $\{Oracle\}_{DB}$ | $S_1 = \{SQL, Java, C, Python, C++, Scala\}_{L}$, $\{.NET\}_{WF}$, $\{TensorFlow, Keras, Torch, Hadoop\}_{DB}$, $\{Kubernetes\}_{P}$, $\{PostgreSQL, MySQL, Oracle\}_{DB}$ |
| $S_2 = \{AWS, Docker\}_{P}$, $\{MongoDB, Cassandra, PostgreSQL, MySQL, SQlite, SQLServer\}_{DB}$ | $S_2 = \{Scala, R, Kotlin, Delphi, Matlab\}_{L}$, $\{TensorFlow, Hadoop, Torch, Keras, Pytorch\}_{DB}$, $\{SharePoint\}_{CT}$, $\{Docker, Kubernetes\}_{P}$ | $S_2 = \{Perl, R, Bash, JavaScript, PowerShell\}_{L}$, $\{jQuery\}_{WF}$, $\{Gitlab, Ansible, Puppet, Chef\}_{CT}$, $\{Linux, Windows\}_{P}$, $\{SQLServer, SQLLite\}_{DB}$ |
| $S_3 = \{Groovy, Abap\}_{L}$, $\{Apache Spark\}_{DB}$, $\{IBM Watson, Google Home, Amazon Echo\}_{P}$, $\{Cassandra, Cosmos DB, Redis, Neo4j, MongoDB, Microsoft SQL Server, Couchbase\}_{DB}$ | $S_3 = \{Flask, Django\}_{WF}$, $\{Pandas\}_{DB}$, $\{Jupyter\}_{CT}$, $\{GitHub, Tfs\}_{CT}$ $\{LAMP\}_{L}$, $\{Apache HBase\}_{DB}$ | $S_3 = \{Kotlin, TypeScript, C\}_{L}$, $\{.NET\}_{WF}$, $\{Visual Studio\}_{DT}$, $\{Jira, Confluence, SVN\}_{CT}$, $\{Android, Xamarin, IOS\}_{P}$ |
| $S_4 = \{Ruby, PowerShell\}_{L}$, $\{Pytorch\}_{DB}$, $\{Visual Studio\}_{CT}$, $\{Gitlab, Puppet, Chef, Ansible, GitHub\}$, $\{Tfs\}_{CT}$, $\{Apache HBase\}_{DB}$ | $S_4 = \{React.js, Node.js\}_{WF}$, $\{Django\}_{WF}$, $\{Flask\}_{CT}$, $\{Serverless, LAMP\}_{L}$, $\{DynamoDB\}_{DB}$ | $S_4 = \{[HTML, Abap, CSS, Groovy]_{L}$, $\{Angular, Spring\}_{WF}$, $\{Apache Spark\}_{DB}$, $\{SharePoint\}_{CT}$, $\{Microsoft Azure, Microsoft Azure 1.0, Prefix\}_{P}$, $\{Redis, Cosmos DB\}_{DB}$ |
| $S_5 = \{Node.js, React.js, Xamarin, Django, Flask\}_{WF}$, $\{Confluence\}_{CT}$, $\{Serverless, LAMP\}_{L}$, $\{DynamoDB\}_{DB}$ | $S_5 = \{[PHP, Perl, Ruby, PowerShell]_{L}$, $\{Drupal\}_{WF}$, $\{Ansible, Puppet, Chef, Gitlab\}_{CT}$, $\{Elasticsearch, SQLServer\}_{DB}$ | $S_5 = \{[Ruby]_{L}$, $\{Jupyter\}_{CT}$, $\{Tfs, Git, Github\}_{CT}$, $\{Docker, AWS\}_{P}$, $\{Apache HBase\}_{DB}$ |
| $S_6 = \{TypeScript\}_{L}$, $\{Angular\}_{WF}$, $\{Git, Jira, Mercurial, SVN\}_{CT}$ | $S_6 = \{Spring, Angular.js\}_{WF}$, $\{MySQL, PostgreSQL\}_{DB}$ | $S_6 = \{VBA\}_{L}$, $\{Windows Phone\}_{P}$ |
| $S_7 = \{Objective-C, Swift, Rust\}_{L}$, $\{Angular.js\}_{WF}$, $\{IBM Watson, Amazon Echo, Google Home\}_{P}$, $\{Neo4j, Couchbase\}_{DB}$ | $S_7 = \{[bash]_{L}$, $\{Windows, Linux\}_{P}$ | $S_7 = \{React.js, Node.js\}_{WF}$ |
| $S_8 = \{[Groovy, Abap]_{L}$, $\{Apache Spark\}_{DB}$, $\{Mint\}_{CT}$, $\{Redis, Cosmos DB\}_{DB}$ | $S_8 = \{[NET Core, ASP.NET Core]_{WF}$, $\{Slack, Slack 1.0\}_{P}$ | $S_8 = \{[Matlab]_{L}$, $\{Cassandra, DynamoDB\}_{DB}$ |
| $S_9 = \{Vue.NET\}_{L}$, $\{ASP.NET, NET\}_{WF}$ | $S_9 = \{Objective – C, Swift\}_{L}$ | $S_9 = \{[Assembly]_{L}$, $\{Express\}_{WF}$ |
| $S_{10} = \{Julia\}_{L}$, $\{MariaDB\}_{DB}$ | $S_{10} = \{Julia\}_{L}$, $\{MariaDB\}_{DB}$ | $S_{10} = \{[Assembly]_{L}$, $\{C#\}_{WF}$ |
| $S_{11} = \{RStudio, IPython\}_{CT}$ | $S_{11} = \{[React.js]_{WF}$, $\{Google Cloud Platform\}_{P}$ | $S_{11} = \{Eclipse\}_{CT}$, $\{Unity 3D\}_{P}$ |
| $S_{12} = \{Microsoft Azure, Microsoft Azure 1.0\}_{P}$ | $S_{12} = \{Microsoft Azure, Microsoft Azure 1.0\}_{P}$ | $S_{12} = \{ React Native\}_{WF}$, $\{SVN\}_{CT}$ |
| $S_{13} = \{Asp.NET Core, .NET Core\}_{WF}$ | $S_{13} = \{[IIS, Android, Xamarin\}_{P}$ | $S_{13} = \{React Native\}_{WF}$, $\{SVN\}_{CT}$ |
| $S_{14} = \{Android, IOS\}_{P}$ | $S_{14} = \{Jira, Confluence\}_{CT}$ | $S_{14} = \{React Native\}_{WF}$, $\{SVN\}_{CT}$ |
| $S_{15} = \{Eclipse\}_{CT}$, $\{Unity 3D\}_{P}$ | $S_{15} = \{React Native\}_{WF}$, $\{SVN\}_{CT}$ | $S_{15} = \{React Native\}_{WF}$, $\{SVN\}_{CT}$ |

**Note**
- L = Languages, WF = Web Frameworks, BD = Big Data/ML, DT = Developer Tools, CT = Collaboration Tools, P = Platforms, DB = Databases
- This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3124909, IEEE Access

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
On the other hand, the multifaceted nature of DA concept comprising a variety of technologies and interdisciplinary skills has led to multiple and contradictory definitions [75]. To mitigate this validity threat, we decided to adjust through an iterated process based on manual trial searches the final set of keywords incorporating terms found in related literature that best describe this technology trend. Secondly, the initial set of candidate job posts was carefully examined to exclude posts that were irrelevant to our study. To mitigate the bias from this action, the exclusion phase was performed by the first and the second author independently and by discussing conflicts. Moreover, the collection, feature extraction, data cleaning and pre-processing processes were implemented through mature packages of Python and R.

The construction of the hard skills lexicon used in the competence detection stage presented in the collection of job openings could also comprise a threat to the internal validity. More specifically, there is always the risk of omitting technologies that are either rarely applied or they are relatively new (e.g. new programming languages or frameworks). Our reliance on the Stack Overflow Developer Survey aimed in constructing a robust reference lexicon that contains hard skills directly mentioned in a primary ICT support hub. We deem this rationale suitable for the targeted purpose of this study i.e. design of a general and representative competency map that will bring up to surface prominent hard skills that are needed in the O&G industry.

Concerning the external validity, a potential threat to the generalization of the results is certainly related to the selection of the job portals used for data collection purposes. To mitigate this threat, we decided to follow a multi-source search strategy on three dedicated job portals. More specifically, we took into consideration domain specific O&G job boards instead of more generic ones (i.e. Indeed, Monster, Glassdoor, LinkedIn etc.) to ensure a high coverage of related job posts. In addition, we followed a data fusion process combining relevant and meaningful information from multiple sources, so as to achieve generalizability of the inferential process than that provided by a single job portal search strategy. Additionally, we have to note that the data collection process was conducted on specific timestamps. Having in mind that the digital transformation is a rapidly evolving journey characterized by incessant changes in prerequisite skills and competencies, the extracted findings may vary during the evolution of this phenomenon. To mitigate this threat, the extraction of data was conducted at two separate time periods covering 2019 and 2020.

VIII. IMPLICATIONS TO STAKEHOLDERS
The aim of this paper was to gain deep insights regarding the demand of the future workforce in the uprising digitalized O&G 4.0 era. More specifically, we investigated the labor market needs by proposing a unified framework mining information from jobs postings found in well-known O&G job boards. The motivation of the current study was the fact that although the leaders and policy-makers have already understood the benefits and effects of the digital disruption, the O&G industry still faces significant challenges, a fact that is also highlighted by annual reports and low maturity indices related to this transition process. In our study, we focused on human capital by analyzing the current demand for digital workers and the determination of job roles and expertise expressed by a set of hard skills required to fulfill a position across emerging technology trends. We believe that the proposed methodology and the main outcomes through its evaluation provide useful implications to several stakeholders and target groups.

Industry leaders and decision makers: The digital transition into the new O&G 4.0 implies the strong compliance of the industry leaders and decision makers to the current labor market needs in order to endorse digital initiatives into their traditional business processes. The findings of the study through the identification of digital job roles, responsibilities and required competencies provide a roadmap for better understanding and knowledge acquisition of the important ingredients driving this transition journey.

Human resource managers and digital agents: The extracted competence maps and the investigation of the interconnected required hard skills across different technology trends can support several human resource management activities such as, the identification of competence gaps of current employees through assessment mechanisms and the establishment of on-the-job training programs. Competence management also plays a vital role for self-organizational evaluation purposes and creation of key performance indicators concerning the current digital maturity level of an organization. Furthermore, the understanding of the digital expertise demand for a particular industrial sector and its associated required technological skills can be the basis for the creation of well-established e-recruitment policies ensuring high-quality and precise descriptions of job roles according to the required needs of a forthcoming job position.

Governments and higher educational institutions: Traditionally, governments and higher educational institutions are considered significant actors for leveraging competence gaps formed by specific labor market needs. Hence, it is more than evident that they could not be hold intact by this radical wave of digitalization. To this regard, there is an imperative need for governments to invest in digital initiatives programs fostering the on-going collaboration between industry and educational institutions. On the other hand, educational institutions should keep track on the emerging technological innovations and their associated job roles and competencies, so as to prepare the future workforce by reshaping or upscaling its current skills. The findings of this study can be exploited towards two directions. First, they provide a roadmap for evaluating their awareness and preparedness related to required competencies driving the industrial 4.0 transformation journey. Second, based on the identified maturity level in adapting technological concepts and accompanied digital
competencies to their educational learning processes, they can initiate or reform, modules, courses and lifelong learning programs in order to train and prepare their students for fulfilling job positions on the demanding O&G 4.0.

Society (Job seekers/Employees): As the driving force that can support the discussed innovations and transformation of the O&G industry to a modern and digitalized entity, job seekers and employees can greatly benefit from the findings of this study by enhancing the desired skillsets pinpointed by the study of interconnected technologies. Consequently, they can seek alternative digital pathways to employment and secure prestigious positions to well-known organizations, shaping their careers and adjusting them to the needs of the digital era.

Research community: Concerning the research and academic community, the present study constitutes to the creation of body of knowledge regarding the overall status of the O&G industry in terms of digitalization and workforce opportunities. It leverages robust methodologies to present the multifaceted aspects of the digital transformation in appealing and easily interpretable ways, while it establishes a condensed benchmark of desired competencies in three leading O&G Industry 4.0 technology trends.

The findings of this study designate that the analysis of the demand for digital workforce and its associated learning competencies is a dynamic phenomenon that heavily depends on evolving technological advances and specific requirements for each job role. Thus, the current study offers a methodological framework that can be used as the basis for further research. An interesting direction for future work concerns the on-going monitoring and updating of skillsets and competency profiles with the aim of capturing the rapid changes of technological advances and their effects on job openings according to labour market needs. Towards this direction, the concept of competence management would be facilitated by the development of real-time web platforms that would provide detection of emerging competencies, job roles and profiles over time and across different regions. Finally, another interesting topic for further investigation is the deployment of machine learning approaches focusing on the automated categorization of job openings and extraction of skillsets according to the special needs of the three O&G sectors (i.e. upstream, downstream and midstream).

REFERENCES
[1] L. D. Xu, E. L. Xu, and L. Li, “Industry 4.0: State of the art and future trends,” International Journal of Production Research, vol. 56, no. 8, pp. 2941–2962, 2018.
[2] A. Inskpen and K. Ramaswamy, “Breaking up global value chains: Evidence from the global oil and gas industry,” Breaking up the Global Value Chain, pp. 55–80, 2017.
[3] M. S. Sanusi and F. Ahmad, “Modelling oil and gas stock returns using multi factor asset pricing model including oil price exposure,” Finance Research Letters, vol. 18, pp. 89–99, 2016.
[4] Y. He and B. Lin, “The impact of natural gas price control in China: A COMPUTABLE general EQUILIBRIUM APPROACH,” Energy Policy, vol. 107, pp. 524–531, 2017.
[5] M. Zhong and M. D. Bazilian, “Contours of the energy transition: Investment by international oil and gas companies in renewable energy,” The Electricity Journal, vol. 31, no. 1, pp. 82–91, 2018.
[73] S. H. Shah and I. Yaqoob, “A survey: Internet of things (iot) technologies, applications and challenges,” 2016 IEEE Smart Energy Grid Engineering (SEGE), 2016.

[74] F. Samie, L. Bauer, and J. Henkel, “Iot technologies for embedded computing,” Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis, 2016.

[75] A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, “Human resources for big DATA

**KONSTANTINOS GEORGIOU** is a PhD student at the Department of Informatics, Aristotle University of Thessaloniki (AUTH). He holds an MSc degree in Data and Web Science (AUTH) and a BSc in Informatics (AUTH). His research interests include software engineering, data analytics and mining open data and software repositories.

**DR. NIKOLAOS MITTAS** is an Assistant Professor in the Department of Chemistry at the International Hellenic University. He received a BSc degree in Mathematics from the University of Crete and the MSc and PhD degrees in Informatics from the Aristotle University of Thessaloniki (AUTH). His current research interests are focused on the application of statistics and data analytics in the area of software engineering, project management and competence management.

**IOANNIS MAMALIKIDIS** holds a BSc degree in Electrical Engineering from the Eastern Macedonia and Thrace Institute of Technology and a MSc degree on Information Systems from the Aristotle University of Thessaloniki (AUTH). His research interests are focused on data analytics and machine learning in the area of software engineering.

**DR. ATHANASIOS MITROPOULOS** holds a BSc degree in Chemistry from the Aristotle University of Thessaloniki and MSc and PhD degrees in Physical Chemistry from the University of Bristol under a research grant from Shell Petroleum. He is currently a Professor in the Department of Chemistry at the International Hellenic University and Dean of the Faculty of Science. He has published more than 100 publications, he is the author of many books, 4 patents and he is running more than 1M€ research programs mainly from private funds.

**DR. LEFTERIS ANGELIS** studied Mathematics and received his Ph.D. degree in Statistics from Aristotle University of Thessaloniki (AUTH). He is currently a Professor at the Department of Informatics of AUTH and head of the STAINS (STAtistics and INformation Systems) research group. His research interests involve statistical methods with applications to information systems and software engineering, especially regarding the research on the human factor, computational and statistics, planning of experiments and methodology in mathematics simulation techniques.

**VOLUME XX, 2017**