



Petroleum Data Governance and Its Impact on Corporate Performance

—The Case of Petroleum Company

Nonye Peter Awurum 

University of Aberdeen, Doha, Qatar

Email: nonyeawurum@gmail.com

How to cite this paper: Awurum, N.P. (2025) Petroleum Data Governance and Its Impact on Corporate Performance. *Open Access Library Journal*, **12**: e12360. <https://doi.org/10.4236/oalib.1112360>

Received: September 23, 2024

Accepted: January 20, 2025

Published: January 23, 2025

Copyright © 2025 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In a general landscape, data governance plays a crucial role in every organization including the energy industry, particularly for petroleum companies that operate in a complex and data-intensive environment. With the upstream, midstream and downstream activities in the petroleum industry, there was a growing need to manage and govern data, which is now perceived as new oil and conceptualized as an asset. While it is acknowledged that data is a valuable corporate asset, many companies fail to exploit it in order to better their performance. Organizations today need to be proactive in their operations and have to make informed business decisions in less time than ever before. This puts pressure on organizations to govern the use of data better. This report examines the significance of data governance within the petroleum sector and its direct impact on corporate performance. By analyzing data modeling, data integration, data quality, security, accessibility, and utilization, we uncover how effective data governance strategies can lead to improved decision-making, operational efficiency, and ultimately, enhanced corporate performance. The report also confirms that poor data governance has a negative impact on corporate performance suggesting that organizations within the petroleum industry need to enhance the quality of data governance in order to realize its full business value and improve business performance.

Subject Areas

Management Organization

Keywords

Data Governance, Corporate Performance, Data Quality, Data Modeling and Compliance with Regulations/Policies/Standards

1. Introduction

Today, the petroleum industry has grown to become data-driven, generating vast amounts of information at various stages, from exploration and production to refining and distribution. Managing this wealth of data is critical to the success of petroleum companies, making data governance a central component of operations in the oil and gas sector. Data governance involves the formal management of data assets, encompassing policies, processes, and controls that ensure data quality, security, and accessibility across the organization. This report explores how petroleum data governance directly influences corporate performance and success.

According to Oto, a standard definition of the term “Data Governance” can be found neither in the research community nor in the practitioners’ community dealing with information systems. However, proposals defining the term agree that Data Governance refers to the allocation of decision-making rights and related duties in the management of data in enterprises [1]. Data Governance specifies a structural framework for decision-making rights and responsibilities regarding the use of data in an enterprise. Data Governance refers to the assignment of decision-making rights with regard to an enterprise’s “data assets” [2].

Organizations today need to be proactive in their operations and have to make informed business decisions in less time than ever before. Organizations face increasing pressure to improve value, accountability, performance, and quality (while reducing risk) to meet the demands of stakeholders, customers, employees, and the government [3]. High-quality data is necessary to meet the organization’s strategic needs and changing organizational requirements. While it is acknowledged that data is a valuable corporate asset, many companies fail to realize its full business value. This has been attributed to poor data governance [4]. An effective data governance program enables the development of formal policies and standards, and ensures oversight over data so that decision-makers may receive accurate and timely information to respond to challenges and opportunities identified above [3].

Research Problem

This study investigates the factors that affect data governance in a petroleum company and also determines the impact of the quality of data governance on the corporate performance of the organization. This study also identifies which of the factors have the greatest impact on data governance.

The oil and gas sectors are faced with an increasing amount of pressure to report a single version of truth. Data governance processes must, therefore, be clearly defined, repeatable and auditable, allowing risks to be quantified and mitigated. Literature shows that the quality of corporate data yields better performance for the firm.

Thus, the quality of data is critical for enterprises in order to be able to meet a variety of business requirements, such as compliance with regulatory and legal provisions, integrated customer management (“360° view on the customer”),

effective and efficient reporting (“single point of truth”), or integrated and automated business processes [1].

Necessity for and Value of Research

In the oil and gas industry, the volume of data is growing rapidly. There are many reasons associated with this rapid growth. One reason is the increase in business activities due to growing competition and costs of mining and processing oil products. This created the need to increase data storage as more fields became cost-effective for mine. The ability to store ever increasing amounts of data introduced the challenge of the organization’s ability to manage, analyze and apply data [3].

Research Questions and Objectives

Research Question

In the oil and gas industry, the volume of data is growing rapidly and there are data challenges facing petroleum industry attributed to this rapid growth. The data challenges appear in the following forms.

- Difficulty of access to the information.
- Need for systems to be better analyzed, managed and standardization of data for a specific query.
- Growing need for technological capabilities which allow time collaboration from anywhere.

Petroleum organizations need to govern their data efficiently and effectively in order to obtain value and insights, increase profitability, and achieve a competitive advantage [3]. Thus, the following research questions are to be answered in order to overcome these challenges.

- *Which factors influence data governance in petroleum sector?*
- *To what extent does the quality of data governance influence the corporate performance of organizations in petroleum sectors?*

Research Objective

To investigate the factors that affect data governance in petroleum organizations and also determine the extent to which the quality of data governance influences corporate performance in the petroleum sectors.

Propositions

Proposition 1: Inadequate compliance with data requirements will negatively affect quality of data governance.

Proposition 2: Inefficiency of data ownership and stewardship negatively affects good data governance.

Proposition 3: Effectiveness of data integration within the organization contributes positively to data governance.

Proposition 4: Inadequacy of data modeling has a negative influence on data governance.

Proposition 5: Effectiveness of data quality contributes positively to data governance.

Proposition 6: When the quality of data governance is poor, it will impact the corporate performance negatively.

2. Literature Review

Data

Data refers to raw and unprocessed facts, observations, measurements, or information collected through various methods, such as surveys, experiments, observations, and sensors. In the context of research and analysis, data serves as the foundation for generating insights, drawing conclusions, and making informed decisions. It can be quantitative or qualitative in nature, encompassing numerical values, textual descriptions, images, audio, video, and more.

The significance of data lies in its potential to provide evidence, support hypotheses, and reveal patterns or trends. Researchers and analysts use data to test their theories, validate research findings, and gain a deeper understanding of complex phenomena. Data-driven decision-making has become a prevalent approach in many fields, including business, healthcare, social sciences, technology and petroleum, as it enables evidence-based strategies and solutions [5]. Data is an intangible asset of great value in an organization. It is the key enabler for efficient processes and the real manifestation of the business as it represents an organization's customers, employees, and suppliers; its activities and transactions; and its outcomes and results [1].

Oil and gas organizations operate in geological complex and remote areas, acquire and process an enormous amount of data throughout the oil and gas production life cycle. Data is acquired to support operations and to monitor operational performance continuously and in real-time to optimize overall organizational performance in production and safety of personnel and the environment [6].

Governance

Governance is a ubiquitous term in the business, and it has different interpretations depending on the perspective of the user. Governance provides a structure for determining business objectives and monitoring business performance to ensure that objectives are accomplished. Governance refers to the approaches that the organization adopts to ensure that strategies are set, monitored, and achieved [7]. In a nutshell, governance empowers the principal to monitor and control the behavior of an agent [3].

Data governance refers to the exercise of authority and control over the management of data. The purpose of data governance is to increase the value of data and minimize data-related costs and risks. Despite data governance gaining in importance in recent years, a holistic view of data governance, which could guide both practitioners and researchers, is missing [8].

Corporate and IT Governance

Corporate governance is defined as the set of processes, customs, policies, laws and institutions influencing the way an organization is administered, controlled and directed. It is largely of interest to the principal stakeholders, that is, board of directors, management and shareholders as it is the discipline which focuses on the proper functioning of management and the goals for which the organization

is governed [2].

IT Governance is a sub-discipline of corporate governance which focuses on the governance, risk and performance management of IT systems. It also includes alignment processes, communication tools and decision-making structures which ensure that IT sustains and extends an organization's strategies and objectives. It is a well-established discipline and can be viewed as an integral part of corporate governance. Tallon defines IT Governance as the management, use, and control of physical IT artifacts (hardware, software, networks) [9]. IT governance is the organizational capacity exercised by the board, executive management, and IT management to control the formulation and implementation of IT strategy to ensure the merging of business and IT. The emergence of the increased interest in IT governance arose after the passing of compliance initiatives, *i.e.*, the Sarbanes-Oxley Act in the United States in 2002 and Basel II in the European Union in 2004, but it has existed from a research standpoint as IT infrastructure, IT business value, IT controls, and project management literature for over two decades [9].

Data Governance

Data governance is the exercise of authority and control over the management of data. It aims at implementing a corporate-wide data agenda maximizing the value of data assets in an organization and managing data-related risks [10]. While data governance used to be nice to have in the past, today, it is taking on a higher level of importance in enterprises and governmental institutions [11]. This is due to some key trends. The amount of data created annually on the whole planet is expected to increase from 4.4 zettabytes in 2013 to 44 zettabytes in 2020. The growing data volumes from diverse sources cause data inconsistencies that need to be identified and addressed before decisions are made based on incorrect data. Companies introduce more self-service reporting and analytics, which create the need for a common understanding of data across the organization. The continuing impact of regulatory requirements such as the General Data Protection Regulation (GDPR) increases the pressure on companies to have a strong handle on what data is stored and where and how the data is being used. Organizations are forced to overcome their challenges regarding inaccurate and incomplete data [8].

Data needs to be governed in order to address data quality issues and for an organization to be able to quantify and measure its data quality. People and tools shape data and determine where it should go. This implies that data governance (DG) is the governance of the people and technology. There are numerous definitions of data governance because this sphere encompasses many things which are required to ensure data quality. Tallon (2013) defined data governance as "the processes, policies, standards, organization, and technologies required to manage and ensure the availability, accessibility, quality, consistency, auditability, and security of data in an organization" [9].

Data Stewardship and Ownership

Data in the organization is always shared, integrated and utilized in inter-organizational ways. Data sharing is defined as distribution of data. Either the well-

-structured or semi structured, among units for further use. Data can be easily replicated and shared across a vast distance. Its value does not decline as the usage increases instead it gains more value [9]. In order for the effective sharing of this data amongst the units of an organization and to avoid conflicts, there should be a clear assignment of the right roles to the right decision areas with the right accountability. This is very critical and important due to access levels of data, namely data enclaves, restricted data and public data. Data enclaves are classified as the most restricted data. Restricted data has lesser access priority, but data policies impose secrecy on it. Lastly, public data has no restrictions and is shared with anyone who requires access to it.

The concept of data stewardship is different from data ownership. Data owners are those individuals or groups in the organization that have a lawful claim towards data and have control capabilities (obtain, create, have access to and the distribution of data). Data owners often belong to the business and not to the IT department of the organization. Data ownership is critical and sensitive due to the fear of data manipulation which may lead to negative consequences for the organization.

The main issue that affects ownership in organizations is data access. Newman (2006) suggested two approaches to address this namely incentives and treating information as a public good. Departments are reluctant to use their best resources (data) for other departments' projects regardless of whether the input will give valuable value to the project or whether it is in the interest of the organization as a whole. They need compensation as being the source of that data. Providing an explicit contract that rewards those who create and maintain data, "ownership" will be the best way to provide incentives [4]. This strengthens the recognition of data ownership rights. Choosing the best incentives which satisfy both principal and agent will be beneficial because organizations will utilize the available technology to its full potential. The subtle intangible costs of low effort will appear as distorted, missing or unusable data which affects quality of data governance negatively.

Alternative approach is to treat data as a public that can be used by different departments of the organization according to principles of data stewardship. Stewardship entities have broad authority to collect, prepare and support the use of data within the organization. Also, to designate certain data uses as being in public interest to be made available to other departments who are able to demonstrate compliance with data stewardship responsibilities [12].

Data Modeling

Conceptual modeling is defined as the process of gathering requirements and clearly documenting a problem domain by use of conceptual models for the purpose of understanding user's requirements and to aid the primary means of communication between the stakeholders such as owners, service providers, business analysts, developers, and users. Conceptual modeling is the cornerstone of many information systems activities not merely to define user requirements, but to also

support development, evaluation, reengineering, acquisition, adaptation, standardization and integration of information systems. Modeling is widely used for database design and management, business process documentation, business process improvement, and software development [13].

Research shows that it is cheaper to remove defects discovered during the requirements stage. Removing the same defect costs on average 3.5 times more during design, 50 times more at the implementation stage, and 170 times more after delivery. This shows that the quality of a conceptual model is of concern because it can affect both efficiency (cost, time) and effectiveness (quality of information systems) of IS development. Studies on the impact of requirements errors showed that in practice even if requirements errors are detected after the analysis stage and are not corrected, it is believed that it is often too expensive or politically unacceptable to correct them [14].

Data Quality

Data plays a vital role in the operation of businesses or enterprises in the information age. Data contributes heavily to the wealth and future success of the enterprises as the businesses produce reports, deliver information, monitor performance, make decisions and achieve competitive advantages based on the data collected. Data volume in organizations is increasing exponentially and data generation has dramatically increased due to rapid changes in Information Technologies. Today's business environment is faced with critical issues of managing and improving the quality of data. When data quality is not dealt with seriously, it can lead to enormous costs of billions of dollars [15]. This shows that firms need to pay more attention to data quality issues, because the business cannot last if it does not have high quality data. Although this research area has been intensively researched there is strong evidence that information quality issues have become increasingly prevalent in today's business practices due to little attention or low priority given to data quality areas as it is overshadowed by issues which are deemed to be important or more pressing [15].

Data Quality is defined in terms of data type and domain, correctness and completeness, uniqueness and referential integrity, consistency across all databases, freshness and timeliness, and business rules conformance. According to Otto, data quality is defined by two aspects, the dependence of observed quality on the user's needs and "fit for purpose", which is the capability to meet the requirements in a specific situation [16]. Data quality consists of six attributes to determine "fit for purpose". These are called data quality dimensions. These dimensions are accuracy, reliability, timeliness, relevance, completeness, currency, and consistency. Consistency determines if a data unit is specified the same throughout the system that is checking violations of semantic rules defined over data items. Accuracy defines how close a data item is to its true value in terms of meaning and "truthfulness". Completeness is measured according to population checks of completeness of columns of a table containing data. Timeliness describes promptness, freshness and frequency of updates of data [8].

Compliance

Companies are required to comply with external regulations and also internal corporate governance policies designed to increase transparency, accountability and to prevent fraudulent activities. Companies must streamline the collection of reporting data to ensure compliance with internal policies (for data security and privacy), external regulations such as Sarbanes-Oxley (SOX) Act, Control Objective for Information and Related Technologies (CobiT), and standards for data exchange (EDI, HL7, SWIFT, etc.). Cobit is a generally accepted framework used by IT auditors to assess SOX compliance. There are concerns about how to handle data when trying to comply with these regulations. Concerns include sensibility in controlling data access and the fact that there is no clean, accurate data for auditors. These concerns can be addressed through information security policies. [13]

Information security policies are defined as the processes and procedures that the employees should follow in order to protect the confidentiality, authenticity and non-repudiation, integrity and availability of information as it is a valuable asset of the organization. Auditors use these policies as guidelines that dictate the rules and regulations of the organization, which in turn govern the security of information [17]. Risks associated with improper information security policy compliance incur huge damages to organizations like corporate liability, loss of credibility, and monetary damage [18].

Although these policies can be clearly defined and detailed, extensive literature shows that the results are not as desirable, because employees seldom comply with information security procedures [19]. This creates a major impediment for the organizations because they need to develop strategies for improving their employees' adherence to information security policies. If an organization can overcome this barrier it can benefit from information security policies. When employees properly adhere to them these are the four outcomes; strategic alignment, value delivery, risk management and performance measurement [20].

Corporate Performance

Organizations are spending lots of money on IT with little to show for it in the output statistics. Although IT investment is associated with superior performance, it is difficult to prove this claim as it is associated with increasing productivity rather than financial impact like return on asset and return on equity. Practitioners should rigorously justify investments in technology and be able to show the gain to get buy-in from top management. Corporate performance measurement gives little attention to business process measurement as the focus is strongly on the traditional functional structure. List & Machaczek emphasize the importance of measurement as measurements are the key. If you cannot measure it, you cannot control it. If you cannot control it, you cannot manage it. If you cannot manage it, you cannot improve it. This clearly shows that measurements are important for control and improvement of the current processes in place [21].

Most previous research when measuring improved organization performance

due to IT investment excessively focuses on financial indicators such as return on investment, return on assets and ratio of expenses to income. Literature on organizational effectiveness showed that organizational performance should not be defined by financial indicators. Rather, it depends on how the organization is viewed. Effective performance of the organization can be measured by the organization's ability to garner scarce resources and effectively turn them into valued outputs. The three main perspectives on organizational performance are: 1) Successful goal accomplishment is the appropriate measure of performance if the organization is viewed as rational, goal-seeking entities, successful; 2) degree of satisfaction of constituents such as employees and customers if the organization is viewed as coalitions of power constituencies; 3) last perspective holds organizations to be entities involved in a bargaining relationship with their surroundings, importing various scarce resources to be returned as valued outputs [22].

When measuring the extent to which data governance affects corporate performance, Kaplan and Norton (2000) identified four critical key perspectives that can be used. Data governance is associated with increasing productivity rather than financial impact [23]. Otto states that the most common business drivers of data governance initiatives are: 1) to ensure compliance; 2) enable decision-making; 3) improve customer satisfaction; 4) increase operational efficiency; 5) support business integration [16].

Data Integration

Data integration is a process of providing a user with a unified view of data that resides across multiple and heterogeneous data sources [24]. Data integration frees the user from knowledge about how data are structured at the sources and how they are reconciled in order to answer queries. Data integration is essential in large enterprises that own a multitude of data sources to produce data sets that can develop and improve cooperation among the units of the organization [25]. Ziegler & Dittrich (2004) give two reasons for data integration. Firstly, an integrated view is created to facilitate information access and reuse through a single information access point. Secondly, data from different sources provide a comprehensive basis for satisfying a certain information need or query. Data integration can be achieved using one of three approaches namely Application Integration (mediation), database federation and data warehousing [26]. Data integration is one of basic activities used to improve the quality of data as it can both reduce data structural and semantic heterogeneity and redundancy, and increase its availability and degree of completeness [24].

Corporate performance

The effective performance of an organization can be measured by the organization's ability to garner scarce resources and effectively turn them into valued outputs. Literature on organizational effectiveness showed that organizational performance should not be defined by financial indicators but rather by how the organization is viewed [22]. Kaplan and Norton (2000) state that the performance measurement called the balance scorecard identified four critical key perspectives

to be included in measuring corporate performance: financial, customer, internal business process, learning, and growth perspectives [23]. The business drivers of data governance initiatives are: 1) ensure compliance; 2) enable decision-making; 3) improve customer satisfaction; 4) increase operational efficiency; 5) support business integration. Therefore, balance scorecard is deemed to be a good measurement of the impact of data governance on corporate performance [1].

Conceptual Framework

A conceptual framework is a representation of the relationship you expect to see between your variables, or the characteristics or properties that you want to study. Conceptual frameworks can be written or visual and are generally developed based on a literature review of existing studies about your topic. A conceptual framework logically describes the relationship among the concepts applicable to the problem under investigation [3].

The researcher's conceptual model in **Figure 1** shows the proposed independent variables that affect data governance, which have an impact on improving corporate performance. The model has seven constructs which were identified from the existing literature; and also shows the relationships between the constructs that were also found in literature. The conceptual model acts as a guideline to organize the measurement, collection and analysis of data [7].

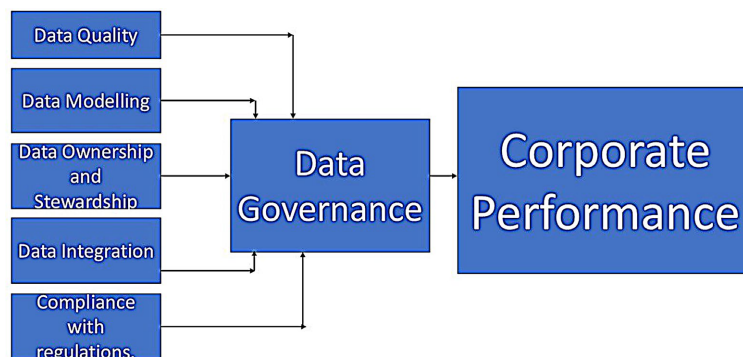


Figure 1. Conceptual model for corporate performance.

3. Methodology

Research methodology refers to the systematic process of planning, conducting, and analyzing research studies. It encompasses the techniques, procedures, and strategies used to gather and interpret data in order to address research questions or hypotheses. Research methodology includes various steps such as selecting a research topic, defining the research problem, choosing appropriate research methods, collecting and analyzing data, and drawing conclusions [7].

Type of Research

This study is explanatory research (quantitative-objective), from the conceptual model; the focus is to find the relationship between independent variables and dependent variables. This study also uses quantitative methods by gathering numerical data through questionnaires.

This study is cross-sectional as this research is for academic purposes (although not to be published in public domain. (K & Saunders, n.d.) says cross-sectional studies often employ the survey strategy to explain how factors are related. This holds true for this research as the researcher aims to investigate how these five constructs (compliance, data ownership, data integration, data modeling and data quality) have an effect on data governance which has an effect on corporate performance [27].

Research Strategy

Njie provides four types of case study designs namely Single case design, holistic; Single case design, embedded; multiple case design, holistic; multiple case design, embedded. The definition of a Case Study is an empirical inquiry.

- Investigates a contemporary phenomenon in depth and within its real-life context.
- The boundaries between phenomenon and context are not clearly evident. This study used a single case design, holistic (single unit of analysis) [12].

Sampling Strategy

The study used non-probability purposive sampling or judgment sampling; this sampling type chooses its sample based on the qualities possessed by the participants. Purposive sampling was appropriate as the research is interested in people who have a rich knowledge of data governance. Data governance incorporates diverse disciplines of data namely data modeling, data quality, data ownership, compliance and data integration. Using purposive sampling ensures that we get participants who have knowledge in all these areas to avoid a situation where participants will answer questions with which they have little knowledge or are not very familiar which increases unreliability. In this research, the researcher aimed to reach 50 or more participants. The population size was 200 that is all staff members who have depth knowledge of data governance. Therefore, the response rate was 50/200 which is the quarter of the population.

Data Collection and Preparation

The data collection technique used affects the results one obtains from the research; therefore, one has to ensure rigor and appropriate methods are used.

There are three methods for conducting research: quantitative, qualitative and mixed methods [12].

As this study adopts a quantitative research paradigm, therefore utilizing a quantitative method was appropriate. Data was collected using quantitative method getting quantitative data with the use of a questionnaire. Likert-scaled items, ranging from 1 (Neither agree nor disagree) to 5 (strongly agree) were used in the scale of an observed variable, accompanied by some questions that require (Yes/No) answers. The first section collected demographic information about participants (e.g., occupation, department team, level of education, etc.) while the second section measured the model constructs (data modeling, data quality, data ownership, compliance and data integration).

Pilot Study

A pilot study was used to minimize bias and test the validity of the research instrument, *i.e.*, a questionnaire. The researcher pre-tested the questionnaire with 12 people from the purposive sample who have deep knowledge (expert) on data governance and people who work with data. This helped me review the questions for relevance and wording. The main objective of this was to ensure clarity and relevance and also to understand the constructs of the variables to be included in the questionnaire.

Data Analysis

Descriptive analysis measures the percentages, measures of central tendency (mean, mode, median) and measures of variability (range, Standard deviation, and variance). Correlation testing will be performed to get a clear view of the relationship between the constructs. Correlation assesses the strength of the relationship between pairs of variables *i.e.*, independent variables. Correlation coefficient ranges from -1 to 1 , where a range from -0.7 to 0.7 shows a weak relationship between the variables assuming calculation on probability less than 0.05 , which is regarded as statistically significant. Linear regression assesses the strength of cause-and-effect relationships between independent variables (*i.e.*, **Figure 1** data quality, data modeling, data ownership, data integration and compliance with regulations) and dependent variables (data governance).

According to an online source, researchers use different data analysis methods depending on whether the data is qualitative or quantitative [28].

Qualitative Data Analysis

Qualitative data is usually in spoken or written information, such as interview transcripts, video and audio recordings, notes, images and text documents. Qualitative data analysis involves identifying common patterns in participants' responses and critically analyzing them to achieve research aims and objectives.

The most commonly used qualitative data analysis methods are:

- Content analysis: This is one of the most common methods used to analyze documented information and is usually used to analyze interviewees' responses.
- Narrative analysis: Researchers use this method to analyze content from several sources, including interviews, observations and surveys. It focuses on using people's stories and experiences to answer research questions.
- Discourse analysis: This method analyzes spoken or written language in its social context and aims to understand how people use language in day-to-day situations.
- Grounded theory: This method uses qualitative data to discover or construct a theory explaining why something happened. It uses a comparative analysis of data from similar cases in different settings to derive explanations.

Quantitative Data Analysis

Quantitative data analysis involves turning numbers into meaningful data by applying rational and critical thinking. Most researchers use analytical software to

assist with quantitative data analysis. The first stage in analyzing quantitative data is validating, editing and coding the data.

Once completed, the data is ready for analysis.

The most commonly used quantitative data analysis methods are:

- Descriptive analysis: This method uses descriptive statistics like mean, median, mode, percentage, frequency and range to find patterns.
- Inferential analysis: This method shows the relationships between multiple variables using correlation, regression and variance analysis.

Access and Research Ethics

The research will be conducted on a petroleum company in Qatar, and this research will deal with sensitive business data since the organization relies on data to make informed decisions and strategies. Permission was asked from the related people who are responsible for data governance initiatives. The introductory letter asking for access describes in brief the purpose of the research, what is likely to be involved in participating, and how the organization can benefit from the study. There are three main organizational concerns, firstly the amount of time or resources that will be involved, secondly sensitivity about the topic and thirdly the confidentiality of the data that would have to be provided and the anonymity of the organization or individual participants.

4. Results

This section presents the results of data analysis and researcher attempted to explore practical and effective data related initiatives (the independent variable from the conceptual framework/model) which contribute to good data governance in order to create better corporate performance in an oil and gas industry.

4.1. Response Rate

The researcher aimed to collect 50 or more responses. The questionnaire was sent to 170 people and 50 were returned, which is 29.4 % response.

4.2. Demographic Analysis

Figure 2 shows that the top three respondents were as follows.

- Specialists = 11 people which is 22% of overall respondents.
- Manager = 9 people which is 18%.
- Analyst = 8 people which is 16%.

4.3. Constructs

Here the researcher presents the results of the items of each independent variables (*i.e.*, Data Quality, Modeling, Ownership, Integrity and Compliance) to see the summary of the respondents' responses.

COBIT 4 defines seven control criteria for information to satisfy business objectives. The respondents were asked to indicate to what extent the organization complies with these controls. The likert-scale with these rating (1 = neither agree

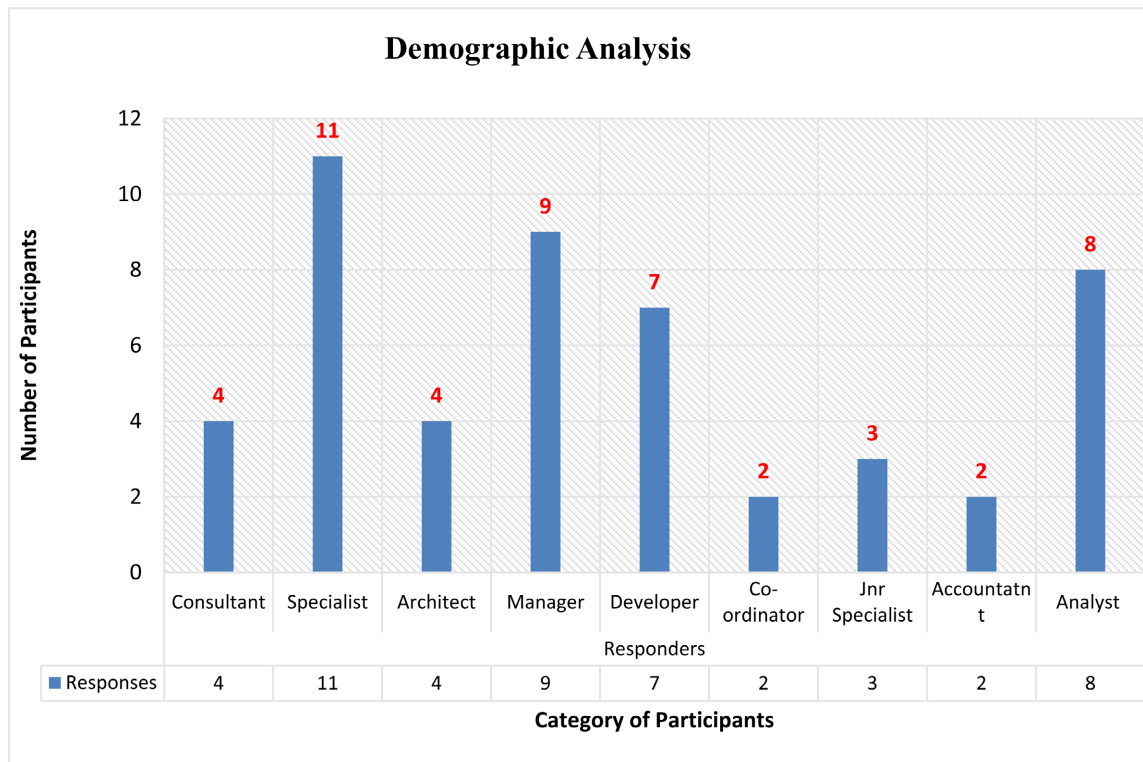


Figure 2. Demographic analysis of participants.

CRL	Description	Valid N	Mean	Std. Dev
CRL1	Effectiveness—information is relevant and pertinent to the processes as well as being delivered in a timely, correct, consistent and usable manner	50	3.66	1.06
CRL3	Confidentiality—Protection of sensitive information from unauthorized person	50	3.78	1.13
CRL4	Integrity—Accuracy and completeness of information as well as to its validity in accordance with business values and expectations	50	3.58	1.14
CRL5	Availability—Information being available when required by the process now and in the future. It also concerns the safeguarding of necessary resources and associated capabilities.	50	3.82	1.06
CRL6	Compliance—complying with the laws, regulations and contractual arrangements, to which the process is subject, i.e., externally imposed business criteria as well as internal policies	50	3.68	1.32
CRL7	Reliability—Appropriate information for management to operate the entity and exercise its fiduciary and governance responsibilities	50	3.46	1.28

Figure 3. Data compliance with data policies and regulations item responses.

nor disagree; 2 = strongly disagree; 3 = disagree; 4 = agree; 5 = strongly agree) was used. Figure 3 above (CRL 1 to 7) shows that the respondents agreed that the organization complies with five of Cobit 4 control criteria for information (Effec-

tiveness, Confidentiality, Integrity, Availability, Compliance) as they have values above 3.5 so they are close to 4 which is agreed on the likert-scale.

From **Figure 4** above (DOS 2 to 6), Respondents disagree that data owners perform their duties regarding data based on the mean values of 3.32, 3.18, 3.14 and 2.74 alongside the Standard Deviations of 1.42, 1.45, 1.43 and 1.45.

DOS	Description	Valid N	Mean	Std.Dev
DOS2	Data owners contribute to definition and management of business metadata.	50	3.32	1.42
DOS3	Data owners determine the transformation rules.	50	3.18	1.45
DOS4	Data stewards support the user community regarding data quality.	50	3.14	1.43
DOS5-6	Data stewards perform exposure or risk identification and verify the data after load.	50	2.74	1.45

Figure 4. Data ownership and stewardship item responses.

DI	Description	Valid N	Mean	Std. Dev
DI2	There is continuous evaluation of existing data integration technology infrastructure and its ability to support data governance practices	50	2.12	1.14
DI3	Data integration lifecycle is followed (Develop and Manage, Access, Discover, Cleanse, Integrate, Deliver, Audit, Monitor and Report)	50	3.1	1.37

Figure 5. Data integration item responses.

From **Figure 5** above (DI), respondents strongly disagree that the organization continuously evaluates the existing data integration technology infrastructure and its ability to support data governance practices.

DQ	Description	Valid N	Mean	Std.Dev
DQ1	Data is Accurate: data item is close to its true value in terms of meaning and truthfulness.	50	3.68	0.87
DQ2	Data is Consistent: data unit is specified the same throughout the organization.	50	3.52	0.91
DQ3	Data is Complete: completeness of columns of a table containing data.	50	3.26	1.05
DQ4	Timeliness of Data: promptness, freshness, and frequency of updates of data.	50	3.36	1.14
DQ5	The organization has Data Quality tools and plans in place.	50	3.44	1.36
DQ7	How often is data cleaning and monitoring done?	50	2.4	1.47

Figure 6. Data quality item responses.

From **Figure 6** (DQ), Four (DQ 1 to DQ4) data quality dimensions were used to measure this construct which are accuracy, completeness, consistency and timeliness. Respondents agreed that the organization complies with two dimensions, which are data is accurate and consistent. But they disagreed that data is complete and timeliness. They also disagreed that the organization has data quality tools and plans in place.

In the case of Data modeling (DM), from **Figure 7**above (DM), respondents

disagreed that data analysts are responsible for developing data models. Data analysts are not the only stakeholders who are responsible for developing data models.

Variable	Description	Valid N	Mean	Std. Dev
DM2	There is a Data model quality management framework which helps in validating the developed data models.	50	3.02	1.45
DM3	Evaluating the quality of a conceptual data model is critical to the successful development of an information system.	50	3.42	1.18
DM4	Data analyst(s) are responsible for developing data models.	50	3.12	1.21

Figure 7. Data modeling item responses.

Variable	Description	Valid N	Mean	Std.Dev
DG1	Organization performs Month-to-month scorecard/KPIs at business unit level.	50	2.62	1.46
DG2	Organization uses Data Quality tools i.e. IBM.	50	2.7	1.59
DG3	How often are data external audits performed?	50	2.04	1.01

Figure 8. Data governance item responses.

Respondents from **Figure 8** above (DG1 to DG3) indicated that the organization has data External audits performed once a year. Respondents disagree that the organization uses Data Quality tools. They also disagreed that the organization performs month-to-month scorecard/KPIs at business unit level for accuracy/quality of specific data entities.

Variable	Description	Valid N	Mean	Std.Dev
P1	Customer-Related Measure: There has been an improvement of customer satisfaction rating as a result of data governance initiatives.	50	2.84	1.58
P2	Financial Measure: There has been a reduction of costs due to improvement in regulatory compliance/reduction of regulatory risk.	50	2.72	1.62
P3	Internal Business Processes Measure: There has been an improvement in internal business processes e.g., quick responses, fewer errors experienced due to data governance initiatives.	50	2.92	1.41
P4	Learning and Growth Measure: There has been a reduction of hours of employee training per employee due to consistent usage of data across the Enterprise.	50	2.5	1.49

Figure 9. Corporate performance item responses.

The balance scorecard in **Figure 9** above, represented by the variables P1, P2, P3 and P4 identified four critical key perspectives to be included in measuring corporate performance namely the financial, customer, internal business process,

learning and growth perspectives. Performance items were around these key perspectives pertaining to data governance initiatives.

4.4. Hypothesis Testing

The following section presents the findings regarding research question one and its supporting prepositions.

Research question 1

Which of the data management practices (factors) have an impact on data governance?

Preposition one, two, three, four and five help in answering the research question one. Multiple regression analysis was carried out to test these hypotheses.

N=50						
	b*	Std.Err	b	Std.Err.	t(44)	p---value
Intercept			---0.87	0.64	---1.35	0.18
AVDOS	---0.05	0.1	---0.05	0.11	---0.44	0.67
AVCRL	0.04	0.1	0.04	0.15	0.30	0.77
AVDQ	0.40	0.1	0.53	0.18	3.00	0.00
AVDM	0.25	0.1	0.27	0.14	1.95	0.05
AVDI	0.24	0.1	0.24	0.13	1.81	0.05
Statistic	Value					
Multiple R	0.70					
Multiple R ²	0.49					
Adjusted R ²	0.43					
F(5,44)	8.49					
p	0.00001					
Std.Err. of Estimate	0.82					

Figure 10. Impact of data management practices on quality data governance..

From **Figure 10** above, The Multiple R (0.70) is the multiple correlation among the five independent variables and the dependent variable, and R Square (0.49) is the variance in the dependent variable accounted for by the five independent variables. The results also showed that data modeling and data integration are significant at 0.05 levels. The F ratio of 8.49 at 5 and 44 degrees of freedom is statistically significant at the 0.00001 level. **Figure 10** shows that Data Quality, Data Modeling and Data Integration are the factors which significantly impact the quality of data governance. So as Data Quality increases, quality of data governance increases; as Data Modeling increases quality of data governance increases; and as Data Integration increases, quality of data governance increases, which are directly proportional.

Data quality has the highest beta value of 0.53 amongst the three significant independent variables, which shows it is a greater predictor with a greater contribution. Data quality variable is significant at the 0.00001 level (*i.e.*, $p < 0.05$).

Research Question 2

What impact does data governance have on the organization's corporate performance?

Proposition 6: When the quality of data governance is poor, it will impact the corporate performance negatively.

N=50	b*	Std.Err.	b	Std.Err.	t(48)	p---value
		of b*		of b		
Intercep			0	0.35	2.09	0.04
AVDG	0.6	0.10	0	0.13	6.64	0.00
Statistic	Value					
Multiple R	0.69					
Multiple R ²	0.48					
Adjusted R ²	0.47					
F(1,48)	44.1					
p	0.00					
Std.Err.	of	1.02				

Figure 11. Impact of quality of data governance on corporate performance.

To test this proposition simple regression was carried out. From **Figure 11** above, R square (0.48) show that 48 percent variance in corporate performance is explained by quality of data governance. F-Ratio of 44.10 at 1 and 48 degrees of freedom is statistically significant at 0.00 level that is p-value ≤ 0.05 . Proposition 6 is supported. The beta value has a positive value indicating that there is a positive relationship between the quality of data governance and corporate performance. This shows that changes in the corporate performance are related to changes in the quality of data governance. A linear regression established quality of data governance could statistically and significantly predict corporate performance. The regression equation was: predicted corporate performance = $0.74 + 0.89 X$ (quality of data governance).

5. Discussion

The primary purpose of this study was to identify and investigate factors that affect data governance in organization and also determine the influence that the quality of data governance has on the corporate performance of the organization.

5.1. Research Question 1

The discussion on research question one consists of a discussion on results of proposition one, two, three, four and five as these were formulated to answer this question.

The results also showed that data modeling and data integration are significant at 0.05 levels. The F ratio of 8.49 at 5 and 44 degrees of freedom is statistically significant at the 0.00001 level. **Figure 11** shows that Data Quality, Data Modeling and Data Integration are the factors which significantly impact the quality of data

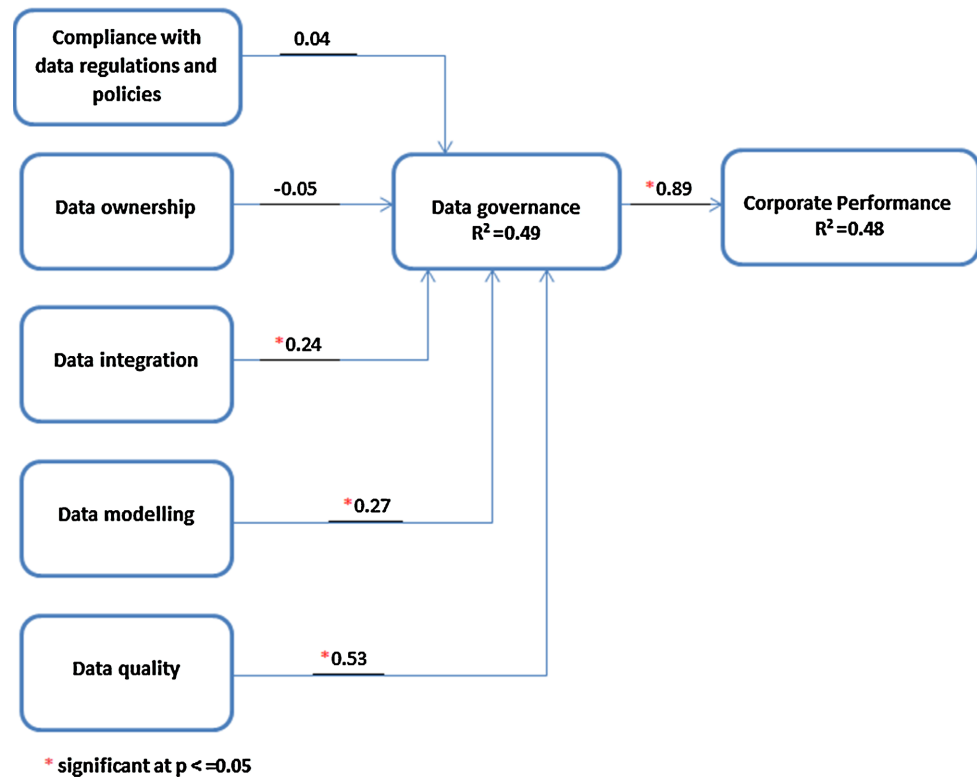


Figure 12. Conceptual model with test results values.

governance. So as Data Quality increases, quality of data governance increases; as Data Modeling increases quality of data governance increases; and as Data Integration increases, quality of data governance increases, which are directly proportional.

From **Figure 12** above, Data quality has the highest beta value of 0.53 amongst the three significant independent variables, which shows it is a greater predictor with a greater contribution. Data quality variable is significant at the 0.00001 level (i.e. $p < 0.05$).

The first proposition states that inadequate compliance with data requirements in organization will negatively affect quality of data governance. When compliance with data policies and regulations was tested against data governance, it did not appear to be statistically significant, implying that compliance with data policies and regulations did not affect quality of data governance. This proposition was not supported because the beta value was not statistically significant; therefore, there was no relationship between the two. This is contrary to the finding of a paper by (Al-Ruithe *et al.*, 2019) which states that data governance is associated with data compliance. Moreover, it was found that complying with the data regulations and policies will increase quality of data governance (Bhansali, 2013).

5.2. Research Question 2

What impact does data governance have on the organization's corporate performance?

The sixth proposition states that: When the quality of data governance is poor

it will impact the corporate performance negatively. We saw that there is a relationship between data governance and corporate performance. This is a strong link since data governance explains 48 % of variance in total. (Tallon *et al.*, 2013) also confirms that there is a link between data governance and firm performance.

5.3. Limitations

The first limitation of this study relates to the selection of participants. Some of the participants did not have depth and broad knowledge of some areas under scrutiny. In sections where the participants had little knowledge, they may have answered the questions by guessing or not answering the question at all.

Another limitation could have been the sample size. Although the sample size of 50 participants is acceptable but can still be regarded as small. This could have been improved by having an earlier start in data collection as they would be more time to survey additional participants.

6. Conclusions and Recommendations

6.1. Conclusion

The results also showed that three of the propositions were statistically significant meaning there was a relationship between independent variables and dependent variables. It showed that there is a relationship between data quality and data governance, data modeling and data governance, data integration and data governance and lastly between data governance and corporate performance.

6.2. Recommendation

This research showed that compliance with data regulations and policies has no effect on data governance. Recommendations to improve this finding would be that the petroleum organization needs to create more awareness and build expertise in particular in compliance regulations and policies around data.

6.3. Future Research

The analysis showed that there was poor data modeling, poor data integration and poor data quality which resulted in poor data governance. It may be enlightening to know how these factor practices are implemented or performed. A research strategy which employs both positivist and interpretivist approaches could provide in-depth knowledge in order to determine areas of improvement. A longitudinal research approach could address the issues of causality, and thereby, provide deeper insight into data governance as this study only confirmed relationships. More research in the same or different organizations with bigger sample could provide a better understanding of this framework.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Otto, B. (2011) Data Governance. *Business & Information Systems Engineering*, **3**, 241-244. <https://doi.org/10.1007/s12599-011-0162-8>
- [2] Bhansali, N. (2013) Data Governance. Auerbach Publications. <https://doi.org/10.1201/b15034>
- [3] Kyobe, M. (2014) The Impact of Data Governance on Corporate Performance: The Case of a Petroleum Company by Zimasa Ndamase (NDMZIM001).
- [4] Newman, M.Z. (2006) From Beats to Arcs: Toward a Poetics of Television Narrative. *The Velvet Light Trap*, **58**, 16-28. <https://doi.org/10.1353/vlt.2006.0033>
- [5] Mayer-Schönberger, V. and Cukier, K. (2023) Big Data: A Revolution That Will Transform How We Live, Work, and Think. Google Books.
- [6] Dissertation, M. and Telukdarie, A. (2017) Optimization of Oil and Gas Production Life Cycle through Data Management by TT MUNYAI.
- [7] Kumar, R. (2018) Research Methodology: A Step-by-Step Guide for Beginners. Sage Publications Ltd.
- [8] Abraham, R., Schneider, J. and Von Brocke, J. (2019) Data Governance: A Conceptual Framework, Structured Review, and Research Agenda. *International Journal of Information Management*, **49**, 424-438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- [9] Tallon, P.P., Ramirez, R.V. and Short, J.E. (2013) The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems*, **30**, 141-178. <https://doi.org/10.2753/mis0742-1222300306>
- [10] DAMA International (2009) The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK Guide).1st Edition, Technics Publications.
- [11] Abu Bakar, A. (2019) The Role of Education in Sustainable Development. *Journal of Environmental Studies*, **45**, 37-50.
- [12] Njie, B. and Asimiran, S. (2014) Case Study as a Choice in Qualitative Methodology. *IOSR Journal of Research & Method in Education*, **4**, 35-40. <https://doi.org/10.9790/7388-04313540>
- [13] Alhassan, I., Sammon, D. and Daly, M. (2016) Data Governance Activities: An Analysis of the Literature. *Journal of Decision Systems*, **25**, 64-75. <https://doi.org/10.1080/12460125.2016.1187397>
- [14] Khatri, V. and Brown, C.V. (2010) Designing Data Governance. *Communications of the ACM*, **53**, 148-152. <https://doi.org/10.1145/1629175.1629210>
- [15] Al-Ruithe, M., Benkhelifa, E. and Hameed, K. (2018) A Systematic Literature Review of Data Governance and Cloud Data Governance. *Personal and Ubiquitous Computing*, **23**, 839-859. <https://doi.org/10.1007/s00779-017-1104-3>
- [16] Otto, B. and Weber, K. (2011) Data Governance. In: *Daten-und Informationsqualität: Auf dem Weg zur Information Excellence*, Springer, 267-287.
- [17] Vroom, C. and Von Solms, R. (2004) Towards Information Security Behavioural Compliance. *Computers & Security*, **23**, 191-198. <https://doi.org/10.1016/j.cose.2004.01.012>
- [18] Bulgurcu, B., Cavusoglu, H. and Benbasat, I. (2010) Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly*, **34**, 523-548. <https://doi.org/10.2307/25750690>
- [19] Herath, T. and Rao, H.R. (2009) Protection Motivation and Deterrence: A Framework for Security Policy Compliance in Organisations. *European Journal of Information Systems*, **18**, 106-125. <https://doi.org/10.1057/ejis.2009.6>

-
- [20] Mitchell Williams, S. (2001) Is Intellectual Capital Performance and Disclosure Practices Related? *Journal of Intellectual Capital*, **2**, 192-203. <https://doi.org/10.1108/14691930110399932>
- [21] List, B. and Machaczek, K. (2004) Towards a Corporate Performance Measurement System. *Proceedings of the 2004 ACM Symposium on Applied Computing*, New York, 14-17 March 2004, 1344-1350. <https://doi.org/10.1145/967900.968172>
- [22] Markus, M.L. and Soh, C. (1995) How IT Creates Business Value: A Process Theory Synthesis. *ICIS 1995 Proceedings*, Whistler, 10-16 September 1995, 29-30.
- [23] Kaplan, R.S. and Norton, D.P. (2000) Having Trouble with Your Strategy? Then Map It. *Harvard Business Review*, **78**, 167-176.
- [24] Magnani, M. and Montesi, D. (2010) A Survey on Uncertainty Management in Data Integration. *Journal of Data and Information Quality*, **2**, 1-33. <https://doi.org/10.1145/1805286.1805291>
- [25] Ziegler, P. and Dittrich, K.R. (2004) Three Decades of Data Intecration—All Problems Solved? In: *IFIP International Federation for Information Processing*, Springer, 3-12. https://doi.org/10.1007/978-1-4020-8157-6_1
- [26] Al Sudairy, M.T. and Vasista, T.G.K. (2011) Semantic Data Integration Approaches for E-Governance. *International journal of Web & Semantic Technology*, **2**, 1-12. <https://doi.org/10.5121/ijwest.2011.2101>
- [27] Saunders, M.N.K. and Rojon, C. (2011) On the Attributes of a Critical Literature Review. *Coaching: An International Journal of Theory, Research and Practice*, **4**, 156-162. <https://doi.org/10.1080/17521882.2011.596485>
- [28] Indeed Editorial Team (2024) What Is Research Methodology? Why It's Important and Types. <https://www.indeed.com/career-advice/career-development/research-methodology>