A Snapshot Analysis of the Data Governance Job Market in The United Kingdom

Abstract

Data governance is a relatively new and emerging field that has gained prominence in the data driven world of today. This study focuses on analysing the current state of the data governance job market in the United Kingdom to provide insights that may guide the career decisions of data governance professionals. The data used for the study focused on “data governance” job postings on LinkedIn. The collected data was analysed to identify the region, city, company, and industries with the highest number of job postings. In addition, the most frequent terms in the job titles and the job descriptions were identified using a bag-of-words approach that highlighted the most relevant words. The study found that England and London had the highest number of postings in country and city while Information and communication industry was found to have the highest number of postings. The advertised job titles were wide and varied and aligned with the job roles identified in literature. Also, the categorisation of the most frequent terms across the top five industries indicated a focus on the data governance “implement” action in the job descriptions. However, a key limitation is that the data was collected from only one source in a single research session which leaves room for further research using multiple sources and deeper text mining techniques.

Key words: data governance, job market analysis, web scraping, text mining, United Kingdom

Chapter One: Introduction

1.0 Background

Data governance refers to the formal execution and enforcement of authority over the management of data and data related assets (Seiner 2014). It has emerged as a vital concern for organisations in recent times (Haneem et al. 2019) which may be attributed to the huge proliferation of data because of technological advancements in the world (Zaki 2019). These technological advancements have led to a marked increase in both reliance and dependence on digital technologies (Van Veldhoven and Vanthienen 2022). A huge volume of data is transmitted often unconsciously and invisibly just by connecting to the internet which has digital records of all activities performed online. Consequently, organisations are gathering vast amounts of information, either through intentional data sharing in return for complimentary online services offered by web-based companies or derived from unsuspecting users who are unaware of the degree and mechanisms of data collection and its commercial utilisation within the digital domain. Furthermore, technological developments in data processing have given businesses the ability to analyse and combine massive quantities of data, discovering insights about individuals that far exceed expectations (Flyverbom, Deibert and Matten 2019), thereby leading to the generation of new types of data and information that needs to be governed (Schneider and Kokshagina 2021; Zaki 2019).

According to the European Commission (2023), there is a prediction of a 530% increase in global data volume from 33 zettabytes in 2018 to 175 zettabytes by 2025. The increasingly huge volume of data from various data sources often results in inconsistent data that must be identified and addressed before decisions are made based on incorrect data (Medeiros, Maçada and Hoppen 2021; Fleckenstein and Fellows 2018). Also, firms are increasingly offering self-service reporting and analytics, which increases the demand for shared data understanding across the organisation (Riggins and Klamm 2017). In addition to this, regulatory requirements such as the General Data Protection Regulation (GDPR) places pressure on organisations to closely monitor where and how their data is governed (Abraham, Schneider and Vom Brocke 2019).

Data governance is majorly about formalising the activities around definition, production, and usage of data in organisations (Seiner 2014). However, there is no universally agreed definition of data governance by both practitioners and academics. This was evident when a survey conducted by the Enterprise Strategy Group (ESG) found that different responses to the question “what does data governance mean to the organisation?” was obtained from the Information technology decision makers (ITDM) and the business team. While ITDM viewed data governance as concerned with ensuring that sensitive data is classified and meets legal and regulatory requirements, the business

team viewed data governance as data quality (Leone and Walker 2022). Conversely, in other research carried out by (Begg and Caira 2012) enterprises referred to data governance as government and regulatory bodies such as the UK data protection act and then stated later that it may also refer to internal governance of business data, managing customer data and providing effective data back up and security. However, in this study, the earlier definition would be adopted.

In recent years, the United Kingdom (UK) has established a complex and robust approach to data governance with several regulations and legislation in place to govern the handling and management of data (The Royal Society 2020). The introduction of the General Data Protection Regulation (GDPR) across Europe in 2018 marked a significant transformation in data governance, compelling organisations to pay close attention to how they govern their data (Voigt & Von dem Bussche 2017). This heightened focus on data management has persisted following Brexit, with the UK instituting its own specific legislation pertaining to data governance, such as the Data Protection Act (2018) and the UK Data Protection & Digital Information Bill (GOV.UK 2023). These regulatory measures have resulted in an increased focus within organisations on conscientiously implementing and adhering to data governance practices, thereby reinforcing the commitment to responsible management of data.

The need to comply with these regulations has been identified as one of the key motivations for organisations to implement data governance (Walsh, McAvoy and Sammon 2022). This motivation along with the increasing complexity of data may have contributed to a demand for data governance roles in the UK. Secondly, organisations that implement data governance would usually implement a data governance program. A data governance program refers to an operating model comprising of people, process, technology, and data that determines and guides how data decisions are taken in an organisation (Ladley 2019). Implementation of a data governance program leads to the creation of new job roles in the organisation that need to be filled (Karkošková 2023). However, organisations may choose to fill the role internally or recruit externally to address their organisational needs and meet strategic objectives (Seiner 2014). Hence, in addition to the growing significance, volume, and regulation of data, the data governance job industry has experienced a corresponding growth, attracting professionals in the field (Harnham 2023). However, there is limited research available analysing the data governance job market.

1.1 Problem Statement

Despite the increasing demand for data governance professionals as a result of the growing volume and regulation of data, minimal research has been conducted on the status of the data governance job market. Rather, previous research has focused on defining data governance frameworks,

organisational motivations for implementing data governance, defining the organisational maturity as well as clarifying data governance activities, models and clarifying the concepts and components of data governance in organisations or in the world globally.

The challenge is that aspiring data governance professionals who wish to enter the UK employment market might not have enough knowledge of how the market operates when making their career decisions. This suggests that additional research should be conducted to understand the makeup of UK data governance job opportunities. In addition, the rising demand for data governance skills also necessitates the need to analyse what the job market for data governance looks like in comparison with the recommendations from various authors both academic and practitioners. Hence, this study seeks to bridge the information gap by investigating and providing insights on the current state of data governance job opportunities in the UK.

An examination of the advertised data governance job roles using both quantitative and qualitative approaches will provide an understanding of the job titles commonly advertised for data governance positions. In addition, it will provide an understanding of the terms frequently used in the job postings. As a result, recruiters, employers, and job seekers will be better informed about the current state of the market and will also know what to expect from a data governance role in the UK.

1.2 Research Aim

The main aim of this research is to conduct a comprehensive analysis of the current state of the UK's employment market for data governance professionals, focusing on an examination of advertised data governance job roles using LinkedIn job postings as a source of information. This analysis seeks to provide an understanding of the current state of the job market for data governance professionals.

1.3 Research Objectives

1. To examine the current trends and characteristics of the data governance job market in the United Kingdom (UK).

2. To identify which industries or sectors in the UK exhibit the highest demand for data governance professionals based on the number of job postings for data governance roles.

3. To explore the common and specific job titles and roles associated with data governance positions in the UK, providing an overview of the data governance job titles offered by organisations in job advertisements.

4. To analyse the most frequent terms used in data governance job postings and compare how they are used across the top five industries.

1.4 Research Questions

1. What are the current trends and characteristics of the data governance job market in the United Kingdom?

2. Which industries or sectors in the UK have the highest demand for data governance roles based on the number of job postings for data governance roles?

3. What is the number of job openings with “data governance” in the job title compared with the job openings without “data governance” in the job title?

4. What are the common/specific job titles and roles associated with data governance positions in the UK?

5. What are the most frequent terms found in the job descriptions for data governance job postings?

1.5 Scope of the Study

The scope of this research focuses on a comprehensive analysis of data governance job roles advertised in the United Kingdom based on a one-time research session. Specifically, the source of data is from the online job portals- LinkedIn.

1.6 Chapter Synopsis

This report comprises of five chapters. Table 1 provides an overview of the chapters.

Chapter One: Introduction	Introduces the research topic and gives an overview of the subject of study as well as stating the objectives of the research
Chapter Two: Literature Review	Reviews existing literature related to the subject including the following: i. What is Data Governance? ii. Data Governance Domains iii. Relevance of data governance to organisations. iv. Organisational motivation(s) for implementing data governance.

	v. Data Governance Roles, responsibilities, and Job titles. vi. Data Governance activities vii. The UK Data Governance Landscape viii. Online Job advertisements in research
Chapter Three: Methodology	Describes the methodology used for the research project and the rationale behind the chosen methodology.
Chapter Four: Findings and Discussion	Discusses the data analysis performed and the findings from the research.
Chapter Five: Conclusion	Provides recommendations and limitations and also concludes the findings from the study.

Table 1: Chapter Synopsis (Source: Author)

Chapter Two: Literature Review

2.0 Introduction

The purpose of the chapter is to provide an overview of existing research on data governance roles and responsibilities in organisations. There is limited research on the subject, thus the literature on the topic includes academic journals; articles and research papers, articles from trade publications and trade practitioners and materials published on the internet that covers the specific topic that form a part of the study.

This chapter discusses the broad concept of data governance including the relevance of data governance to organisations, the organisational motivations for implementing data governance, data governance roles and responsibilities and data governance actions that make up data governance activities. In addition, the UK data governance landscape is discussed alongside the relevance of job postings analysis in research.

2.1 What is Data Governance?

Data governance often means different things to different people and organisations (Seiner 2014). While there is a general notion and understanding of what data governance is and why it is important for organisations to invest in it, there is no universally agreed definition amongst both academics and practitioners (Vial 2023; Gupta and Cannon 2020b; Alhassan, Sammon and Daly 2016; Otto 2011a).

i. The Data Governance Institute provides a concise definition of Data Governance as the exercise of decision-making and authority for data-related matters (The Data Governance Institute 2023a).
ii. The Data Management Association (DAMA) International defines data governance as the exercise of authority and control (e.g., planning, monitoring and enforcement) over the management of data assets (Sebastian-Coleman 2018).
iii. The Royal Society describes data governance as a set of actions and actors that shape data management and utilisation (The Royal Society 2020).
iv. Dataversity describes data governance as a collection of components – data, roles, processes, communications, metrics, and tools – that help organisations formally manage and gain better control over data assets (Knight 2021).
v. On the other hand, Seiner (2014) defines data governance as the formal execution and enforcement of authority over the management of data and data related assets and

emphasizes that data governance is majorly about formalising the activities around definition, production, and usage of data in organisations (Seiner 2014).

It is noteworthy that despite the differing definitions, a common theme is that the definitions involve managing data and suggests that data management is fundamentally important to data governance (Gupta and Cannon 2020b).

There is a common misconception that data governance is the same as data management since the concept of data governance involves treating data as an asset. Nevertheless, it has been argued that data management (DM) and data governance (DG) are not the same (Karkošková 2023; Talend 2023; Gupta and Cannon 2020a; Al-Ruithe, Benkhelifa and Hameed 2019). The Data Management Association (DAMA) provides a definition of Data Management as the development, execution, and supervision of plans, policies, programs, and practices that control, protect, deliver, and enhance the value of data and information assets (Al-Ruithe, Benkhelifa and Hameed 2019). Hence, while data governance deals with formalising the processes and policies around managing data, data management refers to the actual act of managing data by implementing the decisions policies and rules made by data governance (Mahanti 2021). Some experts have opined that the relationship between data governance and data management is symbiotic, with data governance being described as the glue that holds the various aspects of data management together (Mahanti 2021; Seiner 2014). Whereas DAMA holds a view that data governance is a component of data management that is central to all other aspects of data management. This includes data architecture, data modelling and design, data storage and operations, data security, data integration and interoperability, document and content management, reference and master data management, data warehousing and business intelligence, metadata management and data quality management (see Figure 1 below) (Sebastian-Coleman 2018). Each of these components of data management may be the major focus of each advertised role, which may be tied to the organisational motivation for data governance implementation.

Figure 1: DAMA Wheel showing the data management framework (Sebastian-Coleman 2018)

2.2 Data Governance Domains

Data Governance is multifaceted and broad often encompassing various domains. Implementing data governance in organisations would usually focus on two or three of these areas and may determine the job roles advertised. Abraham, Schneider and Vom Brocke (2019) identified the following data decision domains based on the work carried out by Khatri and Brown (2010). These domains are:

1. Data Quality: This refers to the suitability of data for a specific purpose or set of objectives within a given context (Mahanti 2021). Unreliable data can lead to a loss of trust by managers, who may then resort to intuition to make decisions and implement strategies (Redman 2013). The transition from data-based decision making to individual judgement emphasises the necessity of having accurate and reliable data suitable for the specific context (i.e., good quality data). On the other hand, data quality management, which is a field of its own; pertains to the application of quality management techniques to assess, measure, and enhance the fitness of data for diverse uses (Sebastian-Coleman 2018). While distinct from data governance, data quality management is often interwoven with it, as roles in data governance may encompass responsibilities related to data quality, recognizing it as an integral part of overall data management.

2. Data Security: This refers to the maintenance of security standards pertaining to the availability, authenticity, confidentiality, integrity, privacy, and trustworthiness of data. In data governance, it entails conducting risk assessments, establishing data security roles, and defining data security policies, standards, and procedures (Abraham, Schneider and Vom Brocke 2019). Hence, these activities may be a part of data governance responsibilities which may be found in the advertised job postings.

3. Data Architecture: This refers to a blueprint used for managing data assets by integrating people, process, policies, procedures, and technologies to align with the organisational objectives (Hooi-Ten Wong, Maarop and Samy 2020). It defines the standard data terms and elements relevant to the organisation (Sebastian-Coleman 2018). Data governance activities in this domain includes definition of the enterprise data requirements, architectural policies, standards, and guidelines. Data architecture roles such as enterprise data architects are usually responsible for meta-data management as well as defining data standards (Abraham, Schneider and Vom Brocke 2019).
4. Data Storage and Infrastructure: This includes the design, implementation and support of stored data and IT artifacts to maximise its value across the organisation (Sebastian-Coleman 2018). Data governance activities in this domain include assessing application and storage landscape in the organisation, planning software applications storage capacity, definition of standards policies and procedures to guide storage and distribution of data as well as stakeholder trainings on storage utilisation (Abraham, Schneider and Vom Brocke 2019; Tallon, Ramirez and Short 2013; Weber, Otto and Österle 2009).

5. Data Life Cycle: Data lifecycle refers to the approach used for establishing the definition, production, retention, and retirement of data (Khatri and Brown 2010). Data governance activities that may be carried out in this domain include identification of data utilising processes and analysis of the data flow to identify the data storage needs. Additionally, it includes a definition of data retention requirements to meet the various stakeholder needs (Abraham, Schneider and Vom Brocke 2019).

6. Meta Data Management: This refers to defining the content of data to improve interpretability by users (Khatri and Brown 2010). Data governance activities in this domain include defining a metadata strategy and standards as well as specifying the process of developing a metadata repository (Abraham, Schneider and Vom Brocke 2019).

7. Data Stewardship: Although, not identified as a part of the main data governance decision domains by Abraham, Schneider and Vom Brocke (2019), data stewardship is considered by some practitioners as an operational aspect of data governance (Plotkin 2020). The data steward role was identified by Khatri and Brown (2010) as being a potential role in the data principles domain of their framework. Data stewardship is described as a data governance

approach that focuses on formalising accountability for managing data resources on behalf of others for the greater good of the organisation (McGilvray 2021). Data Steward roles would often include the business data steward that represents the business domain and the technical data steward that focuses on the technical areas. Business analysts and data analysts are also described as data stewards in some organisations (Hooi-Ten Wong, Maarop and Samy 2020). However, some practitioners have argued that data stewardship should be more about establishing accountability rather than a specific job role or function (Seiner 2014).

2.3 Relevance of Data Governance to Organisations

Several researchers and practitioners have acknowledged that the complexity and volume of data generated by businesses is growing rapidly due to the current state of the use of digital technologies in the world (Al-Ruithe, Benkhelifa and Hameed 2018; Tallon, Ramirez and Short 2013). While it is anticipated that 175 trillion gigabytes (GB) would be generated annually globally by 2025 (European Commission 2023; Davies 2021), it is predicted that companies will continue to be overwhelmed with data for the foreseeable future. With today's technology, data is like gold - having more of it enhances the growth potential of your business and allows you to generate more revenue (Davies 2021). Since data has been identified as an organisation's greatest asset and source of significant risk, it is crucial that organisations have effective governance in place to manage organisational data effectively (IBM Corporation 2007).

Implementing data governance in organisations is expected to yield the following key benefits: First, data governance enables an increase to direct business contributions (increased sales, customers, market share, etc.) through improvement of processes (Ladley 2019). Secondly, the organisations can also increase efficiency (Koltay 2016; Gregory and Bentall 2012) through better data integration and faster information delivery as it maintains confidentiality, integrity, quality, and availability of customer data (Al-Ruithe, Benkhelifa and Hameed 2019; Gregory and Bentall 2012; Gregory 2011). Thirdly, with the development of data-driven products, data governance can enable the monetization of data (Ladley 2019). Furthermore, it contributes significantly to risk mitigation (IBM Corporation 2007) such as reducing fines, reducing reserves, avoiding market share losses, and reducing risk management costs such as insurance premiums. This risk reduction can be the result of compliance with regulatory standards such as the General Data Protection Regulation (GDPR), improved privacy practices, and improved data quality (Ladley 2019). A knowledge of these expected benefits to the organisation would guide the data governance professional in understanding the needs of the organisation and equip them to articulate how the data governance role aligns with the overall business strategy.

However, it may be suggested that the organisational motivation may influence the focus of the organisation regarding the expected benefits.

2.4 Organisational Motivations for Implementing Data Governance

Data governance is relevant in all organisational contexts, domains or sizes, however the motivation that drives the implementation may differ between organisations (Lis and Otto 2020). Also, there is no one size fits all approach to the implementation of data governance in organisations (Lillie and Eybers 2019; Begg and Caira 2012; Otto 2011a; Hopwood 2008). Hence, the roles and responsibilities defined in organisations may differ based on the motivation for data governance hence its implementation is considered by some academics as being wholly operational (Walsh, McAvoy and Sammon 2022; Gregory 2011).

A study conducted by Walsh, McAvoy and Sammon (2022) which was based on the Khatri and Brown (2010) framework found that Data Principles, Data Access (which includes data privacy and security), and Data Quality are the three main decision areas that drive the implementation of data governance within organisations. Thus, a significant motivation for organisations to implement data governance is the desire to improve compliance which may be both compliance with statutory regulations and/or compliance with internal business processes (Walsh, McAvoy and Sammon 2022; Jang and Kim 2021; Sebastian-Coleman 2018). This motivation may be guided by the organisational goals and intent to improve business efficiency and grow revenue.

Otto (2011b) in his research on morphology of data governance identified data governance goals as being formal or functional. He highlighted the following as formal goals- compliance with legal and regulatory provisions, risk reduction, improving decision making, increasing operational efficiency and support of business integration. The functional goals identified include creation of a data strategy, establishing a data policy, establish data quality control, data stewardship, data architecture management, data lifecycle management, data standard and metadata management (Otto 2011b) These goals may influence the nature of the data governance job role design in organisations.

2.5 Data Governance Roles and Responsibilities

Implementing data governance introduces new roles and responsibilities to an organisation (Karkošková 2023). This is considered a critical aspect when setting up data governance for the first time as the roles need to be clearly defined and given a title (Gupta and Cannon 2020a; Handel 2012).

The common roles identified in formal data governance implementation include Executive Sponsor, Data Governance Council, Chief Steward, Business Data Steward, and Technical Steward (Korhonen et al. 2014; Otto 2011a; Wende 2007). DAMA further adds to this structure by acknowledging that organisations may have multiple layers and proposing additional roles such as Chief data officer, Data governance steering committee, data analysts, data owners, business data stewards or subject matter experts and more to address various levels within the organisation (Sebastian-Coleman 2018). Al-Ruithe, Benkhelifa and Hameed (2018) also extend this by introducing roles like cloud manager, cloud provider member, IT member, and Legal member specifically for cloud data governance. Furthermore, Yulfitri (2016) developed an Operational Model of Data Governance in government and proposed the following roles: data governance board, data governance office, data governance coordinator, data governance working group, business data stewards, operational data stewards and technical data stewards. The limitation on this model, however, is that the scope only covers three data management areas- data governance, data quality management, and metadata management. Hence, data governance in different organisational contexts involves multiple roles and varying requirements which is often tied to the data governance structure set up within the organisation (Seiner 2019).

Gupta and Cannon (2020a) stated that most data governance structures have a minimum of three layers. The Executive/Strategic level is responsible for setting direction, establishing strategies, allocating resources, and giving the final approval for policies and processes. The Tactical level is responsible for implementing strategies, coordinating work, and suggesting solutions to potential process or policy changes while the Operational level focuses on implementing the suggested processes and adhering to approved policies. On the other hand, Korhonen et al. (2014) categorised data governance structures into five layers based on an agile governance model. These layers are strategic steering, strategic implementation, tactical level, operational and or daily level and further mapped to commonly identified data governance roles to each of these levels.

Seiner (2014) in his non-invasive operating model of roles and responsibilities proposed a four-layered pyramid model made up of the executive level (acting as the steering committee), strategic level (data governing council), tactical level (data domain steward, data steward coordinator) and operational level (data stewards). The model has 2 supporting layers made up of data governance partners and the data governance team/data governance manager (Seiner 2014). This model was updated in 2019 to include an additional support layer comprised of the data governance working team whose work is geared towards delivering great data for a specific issue (Seiner 2019) (See Figure 2).

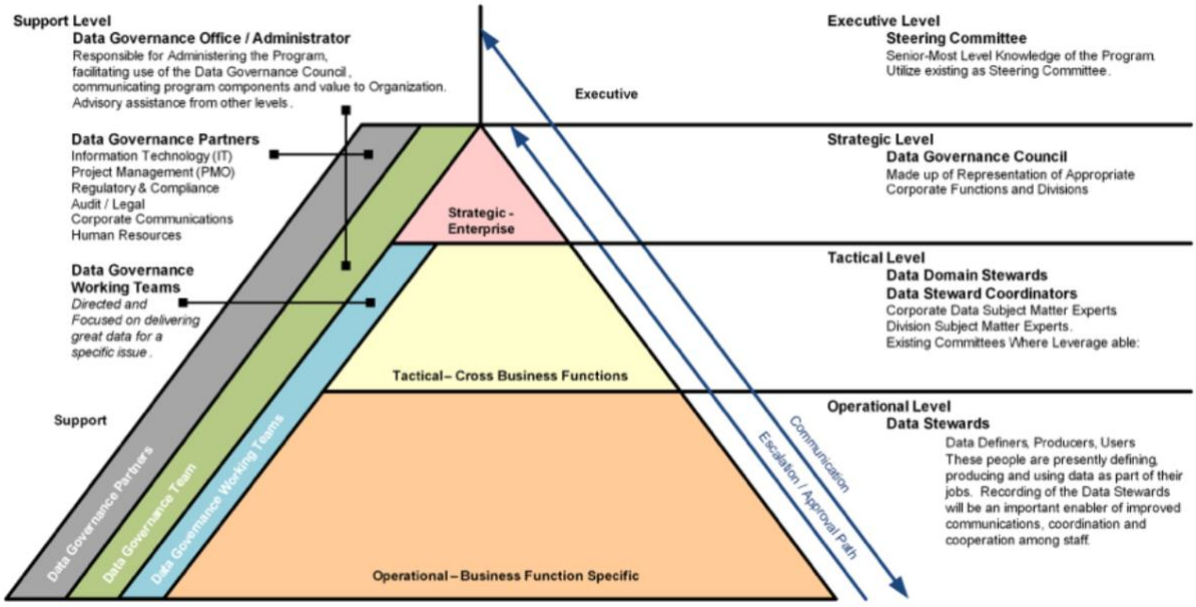


Figure 2:- Data Governance Operating Model of Roles and Responsibilities (Seiner 2019)

Practitioners have suggested that the most critical role for data governance is the Chief Data Officer who has the highest responsibility in an organisation when implementing data governance (Choudhury 2014). The role is an executive role and research has shown that successful data governance in organisations is often driven by top management as it is essential that it is a strategic decision by the organisation (Zhang, Q., Sun and Zhang 2022). Similarly, the data governance office (DGO) is also considered to play a central role in data governance operations first by turning decisions of the strategic and tactical layers into actionable steps and then by facilitating activities and collaborating with the various data stewards and custodians within the organisation to the success of data governance in an organisation (The Data Governance Institute 2023).

Although the data governance community have accepted these roles, they are still considered ambiguous and subject to interpretation within the organisation (Karkošková 2023; Korhonen et al. 2014). Furthermore, while organisations may structure job roles based on various factors, it is relevant to mention that it is possible that these roles may not be advertised with the exact titles referenced.

The data governance roles defined by organisations may be influenced by the following factors: i). The definition of what data governance means to the organisation which may determine the direction taken in data governance implementation (Gupta and Cannon 2020b). ii) Organisational data governance maturity level (Ladley 2019; Seiner 2014) which refers to the current governance processes in place to achieve the strategic and tactical goals and objectives of data governance in the organisation (Gupta and Cannon 2020a). Also, practitioners have indicated that the starting point of the implementation of DG in an organisation is an assessment of the organisational data governance

maturity level (Ladley 2019) which may determine the job roles required. iii) The assignment of locus of accountability for the various decision domains (Khatri and Brown 2010). iv) The question of who drives the DG program- i.e., is it the business or IT? (Cheong and Chang 2007). v) Organisational factors such as number of staff employed by the organisation, number of locations operated in, number of customers, number of systems and processes, number of lines of business as well as number of products may affect the data governance structure and subsequently affect the advertised job roles (Hopwood 2008). vi) Organisational culture and structure (Gupta and Cannon 2020a; Ladley 2019). These factors may subsequently affect the nature of the jobs advertised by organisations and the activities indicated in the job roles.

2.6 Data Governance Activities

According to Alhassan's research (Alhassan, Sammon & Daly 2018; Alhassan, Sammon & Daly 2016), data governance activities are identified as encompassing three core actions, namely “define”, “implement”, and “monitor”. The study found that academic publications emphasised the "define" action, highlighting the significance of establishing data governance foundations such as defining clear roles and responsibilities. On the other hand, practitioner-oriented publications emphasised the "implement" and "monitor" actions, highlighting the necessity of the operational aspects of data governance in practice (Alhassan, Sammon & Daly 2018). It could therefore be said that these operational aspects, which rely on "implement" and "monitor" actions, are vital. Therefore, it is possible that the implement and monitor actions could make up large portions of the duties listed in job advertisements.

However, it is crucial to note that the studies in question mostly rely on systematic literature reviews in a field with limited academic publications, which may bring some subjectivity to the results. Furthermore, the studies draw exclusively on the framework proposed by Khatri and Brown (2010), without making any comparative assessments against other potential frameworks. Hence, it is possible that the methodological decisions taken may have influenced the scope and interpretation of the results.

2.7 The UK Data Governance Landscape

In today’s world, data is considered to be the driving force behind the worlds modern economies fostering innovation in organisations (GOV.UK 2020). The United Kingdom (UK) is regarded as a digital leader, having the largest data market in Europe (i.e., money produced from products or services generated from digital data) (GOV.UK 2022; GOV.UK 2020). For one thing, during the 2010s, the UK's

data economy grew twice as fast as the rest of the economy and, by 2020, accounted for about 4% of the UK's gross domestic product (GDP) (GOV.UK 2020). By 2021, it had the most significant impact among all European Union countries in absolute terms and was estimated to be nearly £125 billion (GOV.UK 2022). Furthermore, in recent research carried out by Digital Planet, the UK was found to exhibit six key identified attributes of data governance having achieved the highest overall score of 89.2% from among the countries assessed for performance (Aaronson, Struett and Zable 2022). In addition, the country demonstrated a wide range of strategies, ethical standards, and human rights commitments, actively engaging in international initiatives related to data governance (Aaronson, Struett and Zable 2022).

The United Kingdom has a complex and robust system for data governance, with several regulations and laws in place to regulate how data is handled and managed. For one thing, recent years have seen a number of policy advances in data governance, including the introduction of new public sector organisations (see Figure 3), the adoption of new rules, and the development of new policy programmes (The Royal Society 2020). In addition, the UK government established an independent data governance panel to improve the use of data across courts and tribunals (GOV.UK 2023b).

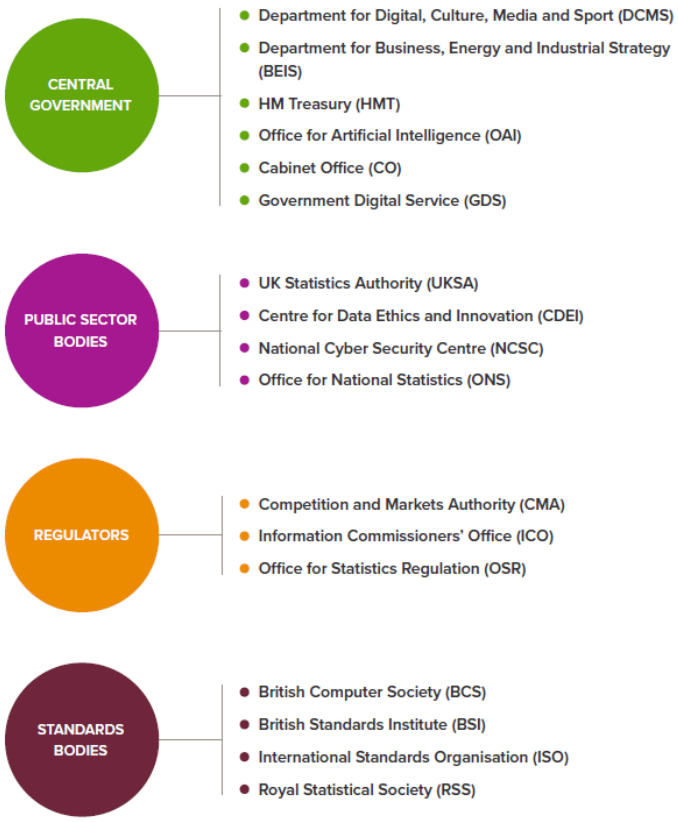


Figure 3: Organisations and structures responsible for governing the use of data in the UK (The Royal Society 2020)

The UK's data governance landscape is shaped by various regulations, policies, and frameworks which include the National Digital Strategy (NDS) released by the UK's Department for Digital, Culture, Media & Sport (DCMS) in September 2020 (GOV.UK 2020), the Data Protection Act (2018) (GOV.UK 2023), the UK Data Protection & Digital Information Bill (GOV.UK 2023), and the Data Sharing Governance Framework (GOV.UK 2022). Equally important is the impact of Brexit and the government's commitment to enhancing data's economic role which highlights the need to develop robust governance structures. In addition, the UK has engaged a combination of previously conducted studies, legislation, and tailored strategies to strengthen its data governance. This makes the UK a rich field for ongoing research in data governance due to the intersection of global trends and local needs.

2.8 Online Job Advertisement in Research

Job advertisements are used in a variety of fields as an important source of knowledge on the evolution of job roles within that field (Pejic-Bach et al. 2020) This is especially true for data governance, which is a relatively new field that is still changing and evolving (Alhassan, Sammon and Daly 2019). Job advertisements offer an overview of the industry's current situation and can aid both researchers and practitioners in comprehending the data governance ecosystem. They can also shed light on the skills and qualifications that businesses seek in a data governance professional. It is therefore a helpful resource for professionals that wish to migrate into the field of data governance.

Job advertisements are often analysed using various methods. However, recent developments have made it easier to analyse job advertisements to extract insights useful for decision making using text mining methodologies. This is because it is now easy to find software for web crawling, extraction, and text analysis (Schedlbauer, Raptis and Ludwig 2021). The ease is further supported by the fact that many organisations now advertise job openings online (Harvard Business Review 2021; Ward-Proud 2014) making it easier to extract the data for analysis. However, despite these advancements, only a few job profiles, such as those for Industry 4.0 and Business Analytics, have been examined using text mining analysis (Schedlbauer, Raptis and Ludwig 2021).

Text mining is a form of text analysis used to extract previously unknown, potential, and practical patterns or knowledge from textual data (Zhang, Y., Chen and Liu 2015). It transforms unstructured textual data into a machine-readable structured form to identify patterns and extract knowledge (Thakur and Kumar 2022). It is becoming increasingly popular among researchers, and it is being used in a variety of fields of study (Thakur and Kumar 2022). Applications of text mining in research include phrase structure recognition, key phrase detection, language identification, authorship relation identification, information retrieval, information clustering, text summarisation and categorisation,

and information extraction (Talib et al. 2016). These varying applications are often found to be relevant to the analysis of job advertisements, however the specific application is dependent on the objectives of the research.

Chapter Three: Methodology

3.0 Introduction

This chapter outlines the methodology used in conducting the research and covers a description of the data collection and analysis procedure. It also discusses the philosophical underpinnings for the methodologies and techniques selected for the study.

3.1 Research Philosophy

The study is based on a pragmatic worldview philosophy that focuses on research problems or questions and uses all approaches available to understand the problem. It is not committed to any one system of philosophy and reality and often applies to mixed methods research where the researchers draw liberally from both qualitative and quantitative assumptions (Creswell and Creswell 2023). This philosophy further allows for the freedom to choose techniques, methods and procedures of research that best meet the intended objectives of the study by focusing on the inquiry or the research questions that need to be answered (Morgan 2017). Using the Dewey's five step model of inquiry as applied to research (see Figure 4), the study starts with an inquiry and further develops until the research is conducted (Morgan 2017).

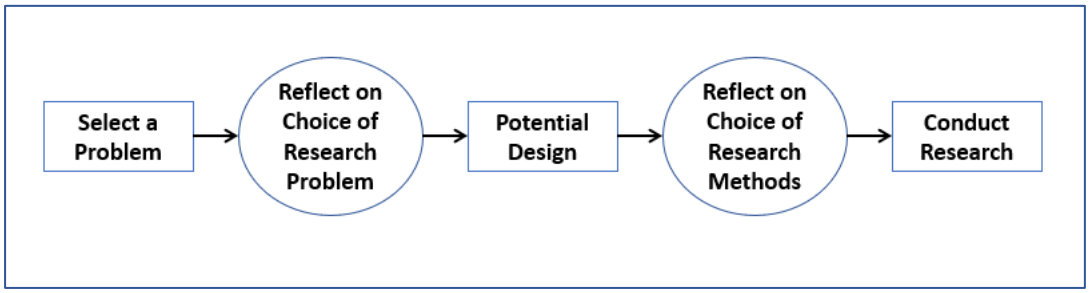


Figure 4: Dewey's five step model of inquiry as applied to research. (Adapted from Morgan (2017))

3.2 Research Approach

The study is based on an inductive approach which involves using the research questions to move from instances gained in the data collection to some form of conclusion, often via comparison with existing concepts or theory. It is exploratory and open ended with data often in the narrative form (Grbich 2013).

The study also takes the form of mixed methods research which can be viewed as a hybrid approach that bridges the gap between qualitative and quantitative research. It aims to consider various viewpoints and perspectives, and gains knowledge from both theory and practice (Johnson,

Onwuegbuzie and Turner 2007). The research design is a qualitative mixed design leaning more towards a qualitative method (see Figure 5). It can also be described as a preliminary quantitative design that starts with quantitative analysis and then further focuses on the qualitative analysis leaning more towards the qualitative analysis (Morgan 2017).

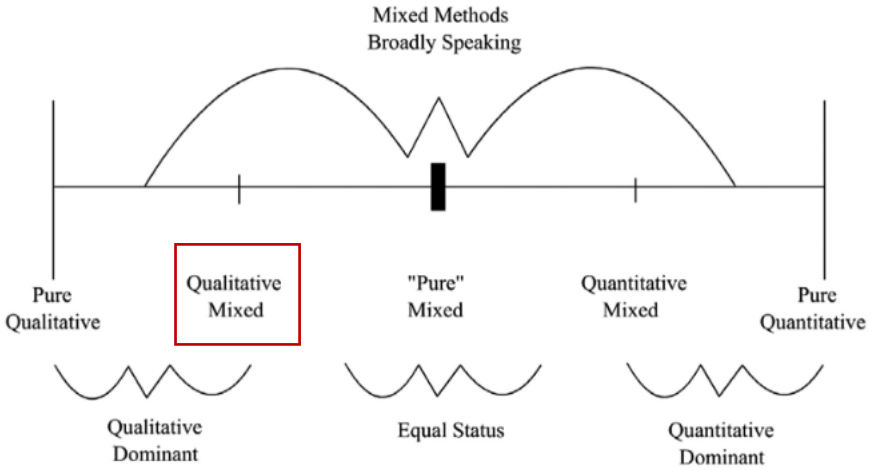


Figure 5: Research Approach, Including Subtypes of Mixed Methods Research (Johnson, Onwuegbuzie and Turner 2007)

Quantitative method focuses on numerical data (Williamson and Johanson 2017). In this study, it was used to analyse the data governance job postings to identify the general characteristics and patterns. This included highlighting the five industries with the highest number of job advertisements. It was also used in the text mining analysis to identify the most relevant job titles and terms used in the job postings. On the other hand, qualitative method was used in the text mining analysis to interpret the context of how the identified terms were used. It has an advantage of being exploratory and is often useful when conducting a study on a topic that not much has been written about in which the researcher seeks to build an understanding of the topic (Creswell and Creswell 2023).

3.3 Research Methodological Approach: CRISP-DM

The methodological approach used in this study is the Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM was created in the 1990s and is widely acknowledged as the most relevant and comprehensive methodology for developing data mining and knowledge discovery projects (Martinez-Plumed et al., 2021; Jaggia et al., 2020). Although it was developed by industry for industry application, it is also considered suitable for research projects and has been used in research projects such as text mining projects (Kurgan and Musilek 2006).

It consists of six major phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment and has the advantage of being iterative allowing movement between the various stages as may be required (see Figure 6) (Saltz 2021; Jaggia et al. 2020).

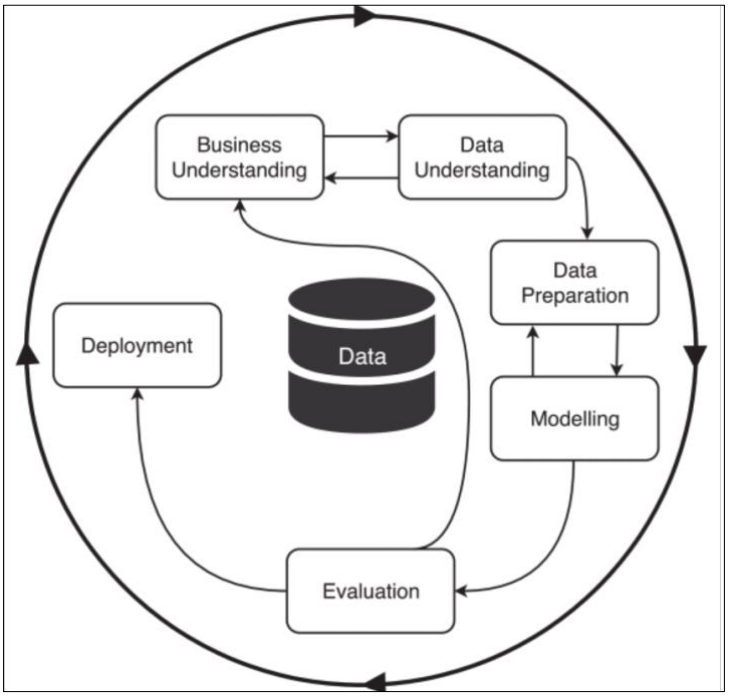


Figure 6: Cross Industry Standard Process for Data Mining (CRISP-DM) (Jaggia et al. 2020)

The six phases can be summarised in the table below:

1.	Business Understanding	The Business Understanding phase is a crucial part of any data mining project and focuses on understanding the objectives and requirements of the project (Saltz 2021)
2.	Data Understanding	The Data understanding phase focuses on identifying, collecting, and analysing data sets that can help with achieving the research aims and objectives. It includes the processes of collecting the data, examining the data to describe it, exploring the data to identify relationships, and verifying data quality (Saltz 2021).
3.	Data Preparation	The data preparation phase involves the steps taken to prepare the final dataset to be used in the modelling tool. It consists of the following steps- select data, clean data, construct data, integrate data and format data (Saltz 2021)

4.	Modelling	This phase is often considered the most intriguing part of data science projects (Saltz 2021)and involves the selection and development of analytics techniques and models often using multiple modelling techniques (Jaggia et al. 2020). It consists of the following steps: selecting the modelling technique, generating test design where applicable, building the model and assessing the model (Saltz 2021).
5.	Evaluation	The evaluation phase involves reviewing and interpreting the results of the analysis considering the business objectives and success criteria described in the business understanding phase (Jaggia et al. 2020). It consists of the following steps of evaluating the results to ascertain if the success criteria were met, reviewing the process to determine if properly accomplished and determining the next steps (Saltz 2021).
6.	Deployment	The deployment phase consists of four tasks: plan deployment, plan monitoring and maintenance, produce final report and review project (Saltz 2021). During this phase the knowledge obtained from the data analysis is converted to a set of recommendations and clearly documented to convey the results of the analysis (Jaggia et al. 2020).

Table 2: Summary of the Cross-Industry Standard Process for Data Mining (CRISP-DM) (Source: Author)

3.3.1 Business Understanding

In this phase, the aims and objectives of the research study earlier stated in Chapter one was defined. Furthermore, requirements and resources needed for the study, potential risks and contingencies were assessed in addition to defining what a picture of success looks like for the study.

3.3.2 Data Understanding

In the data understanding phase of this study, the source of data relevant to analysing the data governance job advertisements was identified. Identifying the source of data includes defining the search term to be used for collecting the data as well as selecting a suitable process to collect the data. In addition, the data was explored to identify relationships and determine the quality of the data.

A. Data Source

The data source used for the study is job advertisements. Job advertisements have traditionally been published in printed newspapers, but in the last two decades, they have increasingly been published online, either on specialized job sites or on social networking sites (Harvard Business Review 2021; Pejic-Bach et al. 2020; Ward-Proud 2014). According to research conducted by the European Centre

for the Development of Vocational Training (CEDEFOP) in 2018, the proportion of job vacancies published online ranges from nearly 50% to nearly 100% in the European Union (EU) countries with the UK being among the top ten (Branka 2018).

Data from LinkedIn was used for this study. The LinkedIn Network has emerged as one of the largest posting sites for job advertisements across various organisations, countries, and roles. Its semi-structured format facilitates text mining analysis, and it also qualifies as a suitable source of data for selection of samples and to conduct text mining analysis (Fennell 2021; Pejic-Bach et al. 2020).

Selection of the job advertisement was conducted in the following manner by focusing on jobs strongly related to data governance in the United Kingdom. The search term "data governance" was used to search for data governance jobs on LinkedIn and the search location was "United Kingdom". Also, the data was collected in a one-time research session between 31st July and August 1st, 2023.

B. Data Collection

The data collection method used was web scraping. The process of web scraping involves retrieving specific data from a web page and converting it into a useful format for various purposes, such as research or business purposes (Macapinlac 2019) . Various web scraping tools, applications and techniques exist for the collection of data from the selected online job boards. However, for this study a low-code tool- Octoparse was used for scraping data from the selected online job board. Octoparse is a user-friendly application for bulk web data extraction and web scraping that delivers advanced features including the ability to extract data in multiple formats such as CSV, Excel, JSON, XML and API (Upwork 2022). The application was selected due to the ease of use and ability to extract unstructured and semi-structured data and convert it into structured data tables for further analysis.

The generic task template- Job details by search_LinkedIn available on Octoparse was used to obtain data from LinkedIn and selection of job advertisement data was carried out by using the search term "data governance" and the search country "The United Kingdom" on the Octoparse application (See Figure 7). The web scraping results was then exported and saved in .xlsx format for processing and analysis.

The number of jobs found on LinkedIn containing the search term- "data governance" was 2,439 jobs, however, 1,439 of those jobs were duplicates and was excluded from the exported data. Hence the exported data contained 1000 job adverts.

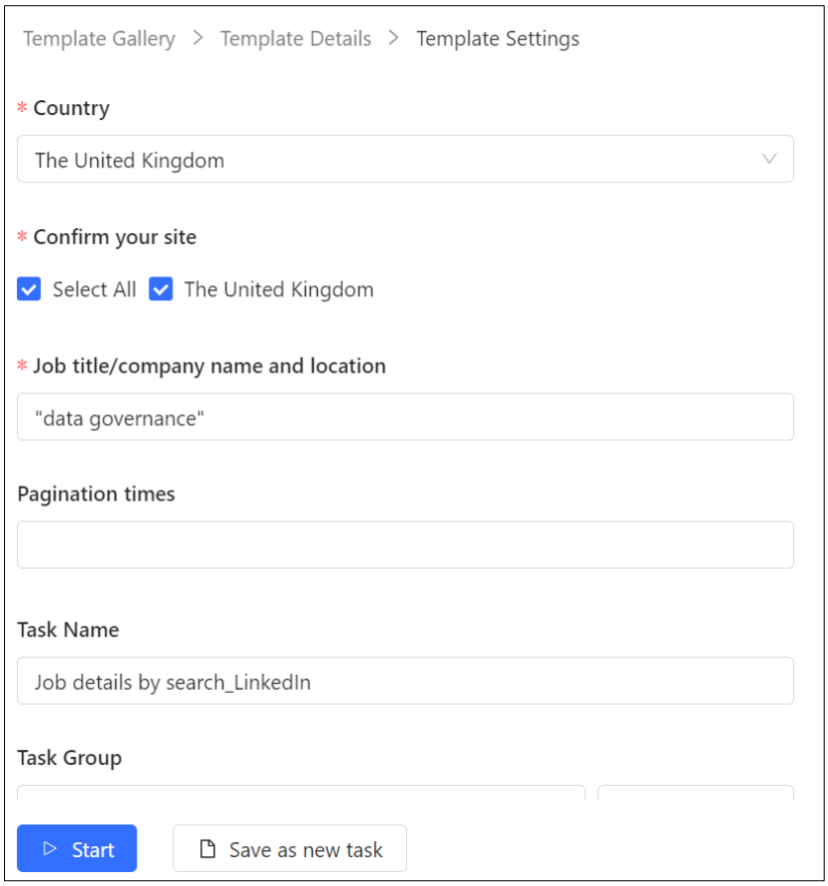


Figure 7: Octoparse Webscraping Details showing LinkedIn search terms (Source: Author's work)

C. Data Exploration

The scraped LinkedIn job advertisements data are semi-structured and includes the following fields: keyword, Location, Job title, Job_link, Company, Company_link, Job location, Job post time, applicant count, Job description, Seniority level, employment type, Job function, Industry and Person hiring (See Table 3).

Field	**Type of field**	**Note**
Keyword	Structured	Free style text
Location	Structured	Free style text
Job Title	Unstructured	Free style text
Job link	Structured	Free style text
Company	Unstructured	Free style text
Company link	Structured	Free style text
Job location	Unstructured	Free style text containing city, region and country
Job-post time	Unstructured	Free style text
Applicant count	Unstructured	Free style text

Job description	Unstructured	Free style text
Seniority level	Structured	Predefined classification
Employment type	Structured	Predefined classification
Job function	Structured	Predefined classification
Industry	Structured	Predefined classification
Person hiring	Unstructured	Free style text

Table 3: Structure of LinkedIn Advertisements Data (Source: Author)

As is visible from the table, most of the fields have free style text input, hence for the purposes of this study, the following fields were used:

- Job title
- Company
- Job location
- Job description
- Seniority level
- Employment type
- Job function
- Industry

3.3.3 Data Preparation

In the data preparation phase, the data was prepared for analysis. Two forms of analysis were carried out in this study- descriptive analysis and text mining analysis, hence data preparation steps would be discussed for both forms of analysis.

The data was considered to be quite clean with about 2% of missing data. Initial data cleaning was carried out using the Tableau Prep tool, although some actions were also carried out manually.

Job location consisted of city, region, and country. This was split into separate fields for Job city and the Job region (England, Scotland, Wales, Northern Ireland). The field for country which was United Kingdom was excluded from the data however where Job country was England, Job city was assumed to be London where data was unavailable. This assumption was made for 13 records.

Industry, Job City, Employment type and Seniority level collectively had 55 missing records and was excluded from the data set bringing the total number of records in the dataset to 945. A further classification was conducted manually on industry using a broad classification that aligns with the UK Standard Industrial Classification (SIC) 2007 classification (See appendix I).

Employment type consists of the predefined values Full-time, Part-time, Contract, Temporary, Volunteer, Internship and was recategorized into Full-time, Contract and Others.

Two fields were created manually- Job title has “data governance” (Y/N) and Number of occurrences of “data governance” in job description.

The prepared data was saved for further analysis and preparation for text mining analysis.

Text Preprocessing

Text mining is a form of data mining used to analyse structured or semi-structured text documents. It can be used in research to identify trends by categorising relevant words and relationships in order to draw conclusions (Pejic-Bach et al. 2020). In this study, text mining was used to analyse the Job title and Job description fields using the RapidMiner Studio version 10.1.003. Duplicate job descriptions were removed leaving 802 records that was used for the text mining analysis.

The data preparation phase of a text mining analysis is referred to as Text preprocessing. The following text preprocessing actions was carried out on RapidMiner (See Figure 8):

i. Extracting content to remove HTML tags.
ii. Tokenization which is the process of dividing text and separating into words or tokens (Shmueli et al. 2023).
iii. Transform cases to lower case.
iv. Filter Stopwords (English): The stopwords used in Rapid miner’s text processing extension was filtered out in addition to own list of stopwords generated to remove words that are not relevant to the analysis (See appendix 2). Stopwords refer to frequently occurring terms that do not provide any additional information such as the, but, best, if, how, less, more, is etc. (Shmueli et al., 2023; Upadhyay et al., 2023).
v. Filter token by length was used to remove tokens less than 4 characters and a maximum of 25 characters.

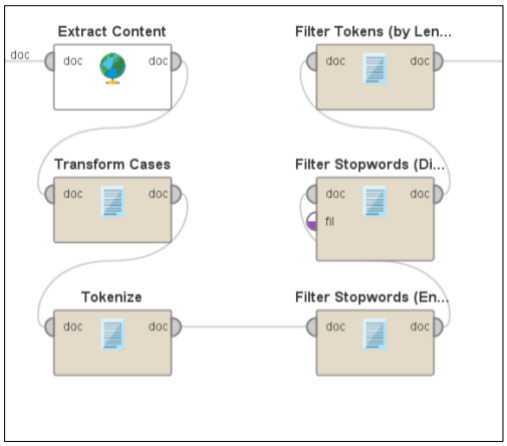


Figure 8: Text preprocessing (Process Documents) on RapidMiner (Source: Author's work)

Key decisions taken in this phase include:

1. Pruning: In this study tokens that appear in less than 10% of the total documents were excluded as well as tokens that appear in more than 95% of the total documents. This process is used to eliminate noise so that the key tokens may be focused on for meaning extraction.
2. Term Frequency- Inverse Document Frequency (TF-IDF): is a combination of Term frequency and Inverse documents frequency (IDF) and is used to rank the importance of a term within a document (Kotu and Deshpande 2014). In this study, TF-IDF was selected on the "Process documents" operator in building the term documents matrix on RapidMiner. Selecting TFIDF removed words not relevant to the document and subsequently generated a dictionary of only relevant words (Upadhyay et al. 2023). It can be calculated using:
    a. Term frequency computes how frequently a term appears in a document and is calculated as:

$$TF(t) = \frac{\text{Number of times term t appears in a document}}{\text{Total number of terms in the document}}$$

    b. Inverse document Frequency (IDF) which is calculated as:

$$IDF(t) = \log\left(\frac{\text{Total number of documents}}{\text{Number of documents containing term t}}\right)$$

    c. TFI-DF is the product of TF(t) and IDF (t):

$$TF\text{-}IDF = TF(t) \times IDF(t)$$

(Shmueli et al. 2023)

3. Lemmatisation: This refers to reducing the words to its base form as may be found in the dictionary (Ignatow and Mihalcea 2018). The advantage of lemmatization is that it is readable and understandable by humans and reduces the number of words to improve the analysis. A disadvantage, however, is that the context of the word used may be lost in the process. In this study, lemmatisation was not carried out.

3.3.4 Modelling

In this phase of the study, data analysis was conducted using descriptive analysis and text mining analysis.

A. Descriptive Analysis

Descriptive analysis was carried out using data visualization techniques to analyse the job postings based on the job role to understand the industry and regional variations, identify the regions and the industries with the highest demand for data governance roles. Additionally, the relationship between the seniority level and contract type was investigated, the top companies advertising for the data governance roles were also identified.

The analysis was used to show the patterns and variations of the data obtained from the job postings data.

B. Text Mining Analysis

In this phase the bag of words approach was used to build a term-document matrix which was further reviewed to draw conclusions. This approach is similar to the approach used by Gupta and Cannon (2020b) in the paper "A Review of Data Governance Definitions and Emerging Perspectives" in which the most frequent terms was analysed. The bag of words approach simply handles each document (record) as a collection of words in which the order, grammar and syntax do not matter (Shmueli et al. 2023).

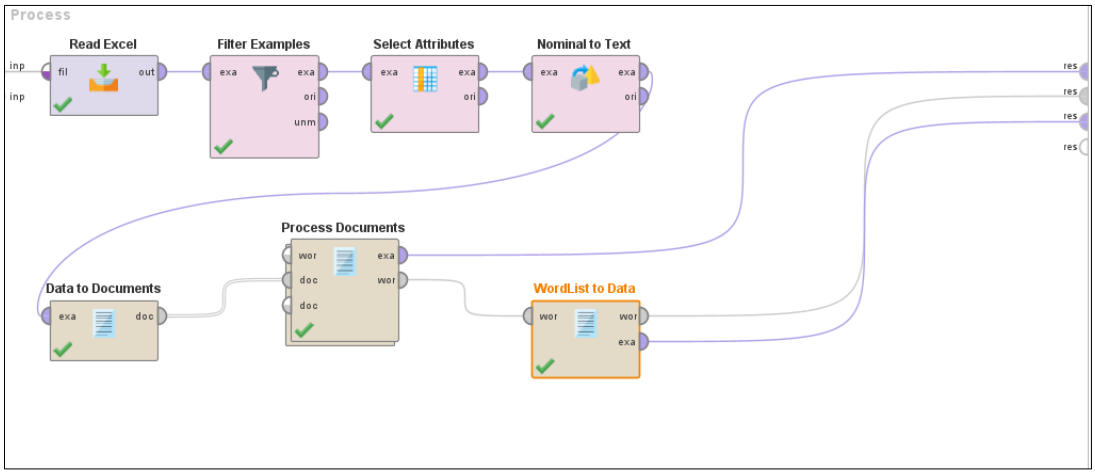


Figure 9: Term Document Matrix Creation Model on RapidMiner (Source: Author's work)

Step 1: The model (Figure 9) was created and used to generate a list of the terms in the corpus based on the grouping of jobs that have the term “data governance” in the job description and those that do not have the term in the job description field.

Step 2: The top twenty most frequently occurring terms were extracted from the job description field across the earlier identified top 5 industries with the highest number of data governance job postings. These terms were analysed and categorised into the data governance actions: “define”, “implement”, and “monitor” for further analysis.

Step 3: Generate a word cloud of the top 250 terms in the corpus for the job title field.

3.3.5 Evaluation

In this phase an analysis of the bag of words and term document matrix created in the modelling phase was carried out and the results evaluated to ascertain its relevance and possible relationship to literature.

3.3.6 Deployment

The deployment phase of this study includes the production of a final written output on the outcome of the study that contains the findings and conclusions.

3.4 Ethical Considerations

From an ethical perspective, selecting appropriate guidelines for text mining research can be challenging, as a comprehensive set of rules is unlikely to address all issues regarding online research. In large part, this can be attributed to the fact that online communities and platforms are so different and unique that only a handful of standards can be designed to satisfy all the specific needs of these communities and platforms (Ignatow and Mihalcea 2018). However, the data source for this study is web data from the public domain with no individual details captured that may be a cause for concern.

3.5 Limitations

One of the possible methodological limitations of the study is the too short period of data collection, however, the study focuses on a snapshot analysis and some researchers have previously also collected data for a short period (Pejic-Bach et al. 2020). In addition, the data was collected from only one job posting site although there are many job posting sites in the UK in addition to companies posting jobs directly on their website.

Secondly, the CRISP-DM methodology is better suited to industry application as part of an ongoing business process that includes a deployment of the improved process and iterations until the business goal is achieved. Hence, when used in research, some of the phases may not be fully applicable within the context described, however it is generally acknowledged that it is a suitable methodology for data/text mining projects.

Chapter Four: Findings and Discussion

4.0 Introduction

This chapter presents the findings and results of the analysis of jobs data collected. It also includes a discussion of the results. The results of the descriptive analysis are presented first, followed by the analysis of the job titles and the most frequent terms.

4.1 Descriptive analysis

This section provides a descriptive analysis of the data governance job market in the United Kingdom.

4.1.1 Job Posting by Country

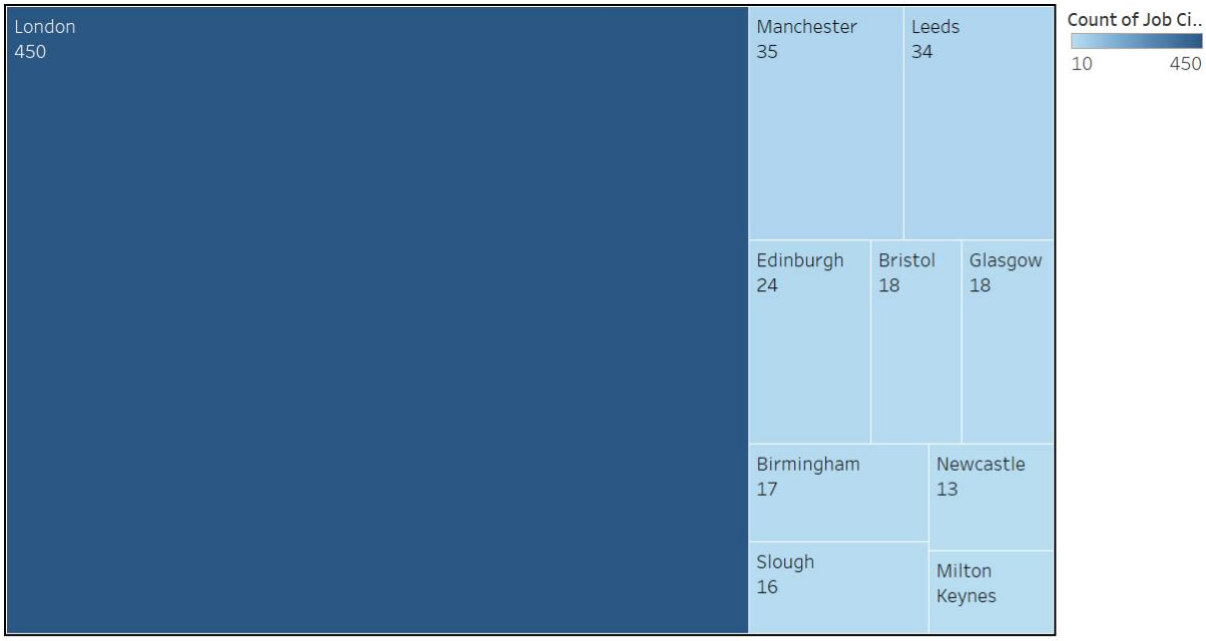
The data on the number of data governance related jobs in table 3 showed that there are more job postings advertised in England than in the other parts of the United Kingdom. England accounted for 93% of the total data governance job postings on LinkedIn while Northern Ireland and Wales accounted for 1% of the job postings each, although Northern Ireland was the least. Therefore, data governance professionals seeking jobs in the UK should focus on jobs located in England.

Country	Number of Job postings	Percentage of Job postings
England	875	92%
Scotland	55	6%
Wales	9	1%
Northern Ireland	6	1%

Table 4: Job Postings by Country (Source: Author)

4.1.2 Jobs by City (Top 10)

As can be seen from Figure 10 below that shows the top ten cities where jobs are posted, London was identified has having the highest number of Job postings. Manchester and Leeds, while not competing with London, also offers some opportunities. For professionals whose focus is Scotland, the two cities- Edinburgh and Glasgow may offer opportunities. However, aspiring data governance professionals should focus their job search in London for optimal results.



Job Cityandcount of Job City. Colour shows count of Job City. Size shows count of Job City. The marks are labelled by Job Cityandcount of Job City.The view is filtered on Job City, which has multiple members selected.

Figure 10: Jobs by City (Top 10) (Source: Author)

4.1.3 Jobs by Industry

Figure 11 shows that the information and Communication Industry has the highest number of job postings and was closely followed by Administrative and support services, while the agriculture, forestry and fishing industry as well as the accommodation and food services industry accounted for the least number of job postings.

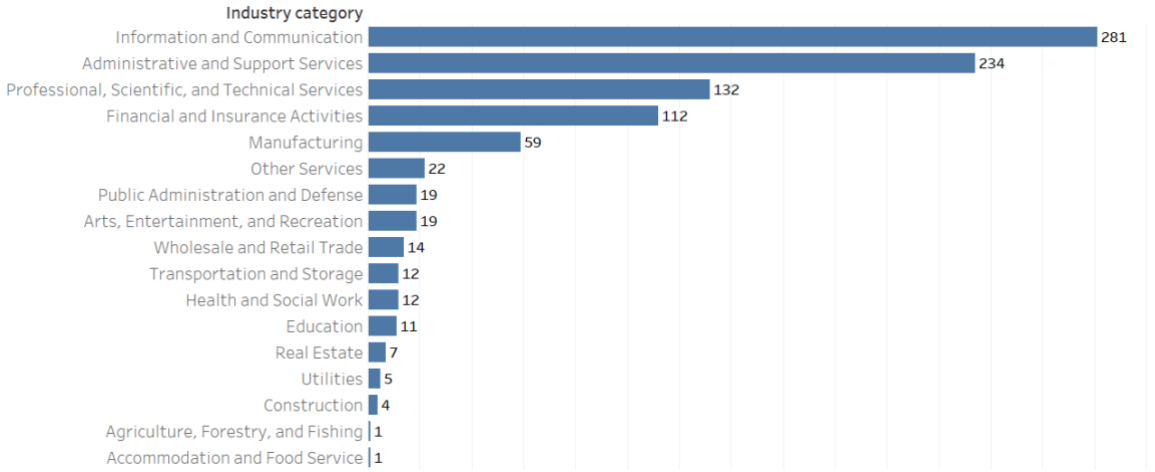


Figure 11: Jobs by Industry (Source: Author)

4.1.4 Jobs by Employment Type

An analysis of the job postings revealed that most of the advertised data governance roles are full time roles (See Figure 12). This may suggest that aspiring data governance professionals should expect to find more full-time roles when searching for data governance roles.

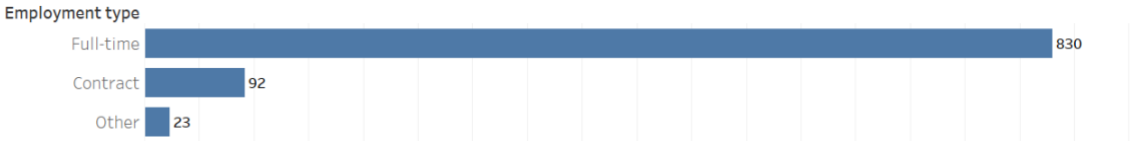


Figure 12: Jobs by Employment Type (Source: Author)

4.1.5 Jobs by Seniority Level

The data revealed that the most common seniority level for the data governance roles advertised were Mid-Senior level while the least was the Executive level roles (see figure 11). Thus, this may imply that aspiring data governance professionals at the mid-Senior level may find more available roles advertised than Entry level, director, associate, or executive levels.

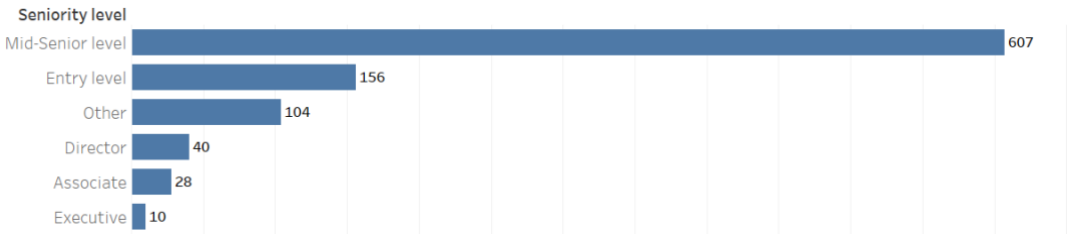


Figure 13: Jobs by Seniority Level (Source: Author)

4.1.6 Jobs by Company (Top 10)

An analysis of the companies with the highest number of job postings was also carried out. This revealed that the company with the highest number of data governance job postings at the time of research was Nigel Frank International (See figure 14). However, it was observed that the companies with the highest number of job postings are recruitment/consulting companies that assist companies to recruit their staff. This suggests that aspiring data governance professionals may look to recruitment firms to apply for jobs rather than focussing on only the company's website details.

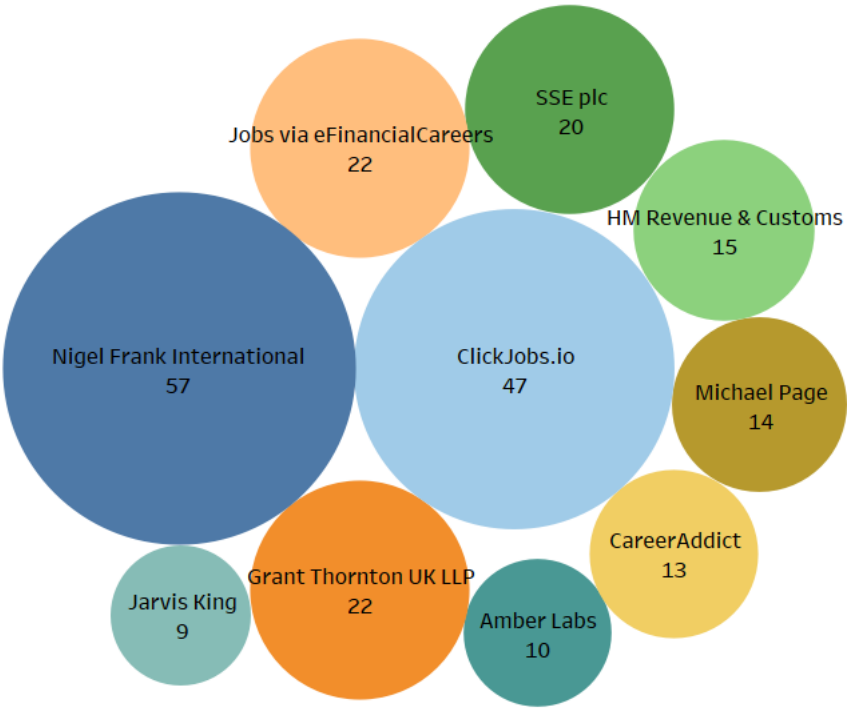


Figure 14: Jobs by Company (Source: Author)

4.1.7 Comparison of Seniority level vs employment type by Industry Top 5

Based on the analysis of jobs by industry (See figure 11), the following industries have been identified as having the highest number of data governance roles advertised from the data collected: Information and Communication, Administrative and Support services, Professional Scientific and support services, Financial and insurance activities and the manufacturing industries. Further analysis by industry will focus on these top 5 industries.

A. Information and communication

Almost all the advertised roles in the Information and Communication industry (see Figure 15), especially those in the Mid- Senior level are full time, which suggests the demand for experienced professionals who are interested in long-term positions. While there is a mix of various employment types across the various levels, only full-time roles were available for the Executive level which is indicative of the level of commitment sought at that level.

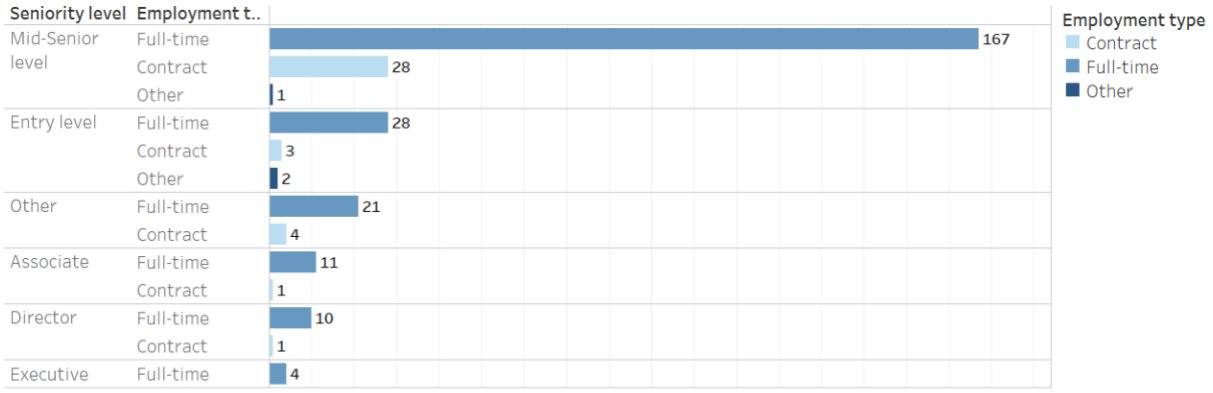


Figure 15: Information and Communication Industry Seniority Level Vs Employment type (Source: Author)

B. Administrative and Support Services

The advertised roles in the administrative and support services industry (see figure 16) are similar to the information and communication industry as the majority of the roles are full time, mid senior level roles. However, there were more entry level roles in this industry when compared to the other industries. This may suggest that entry level data governance professionals should look to the administrative and support services industry to find data governance job roles.

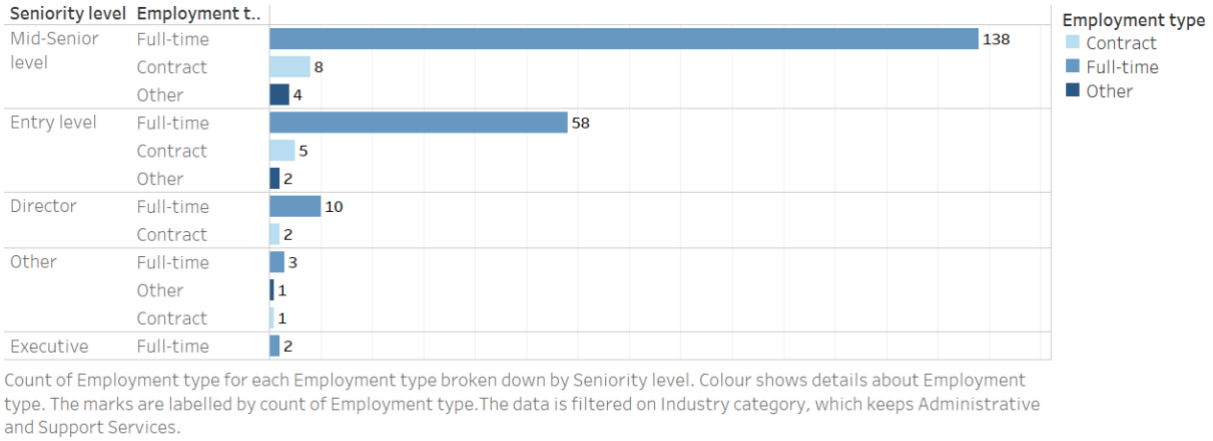


Figure 16: Administrative and Support Services Industry Seniority Level Vs Employment Type (Source: Author)

C. Professional Scientific and support services

The professional, scientific and support services industry is characterised primarily by full time roles being advertised across most seniority levels. The mid-senior level is the only level with multiple employment types (see figure 17). This may indicate that professionals interested in the industry

should focus on full-time roles as other type of roles may not be readily available across the various seniority levels.

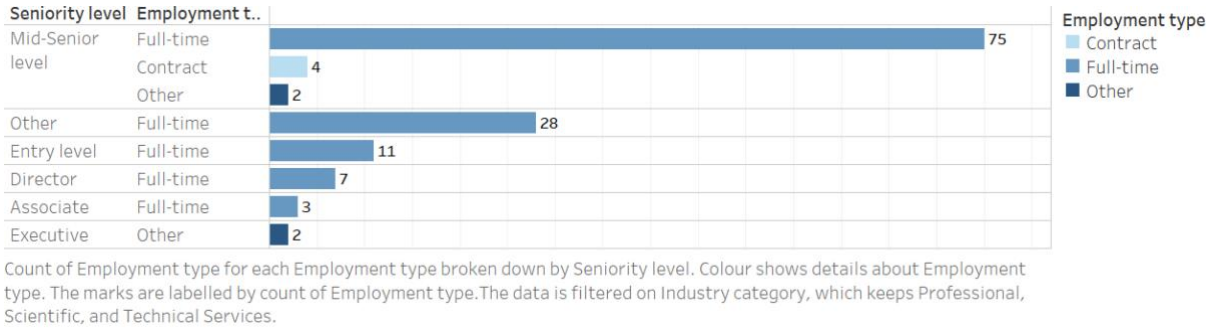


Figure 17: Professional Scientific and Support Services Seniority Level Vs Employment Type (Source: Author)

D. Financial and insurance activities

The advertised roles in the financial and insurance activities industry is characterised with a mix of seniority levels and employment types (see figure 18). The mid-senior level is the predominant level along with the full-time roles. Hence, while there is an availability of varying levels and employment type in this industry, it may be suggested that aspiring data governance professionals would have more opportunities in mid-senior level roles, other and entry level roles compared with the executive and director roles.

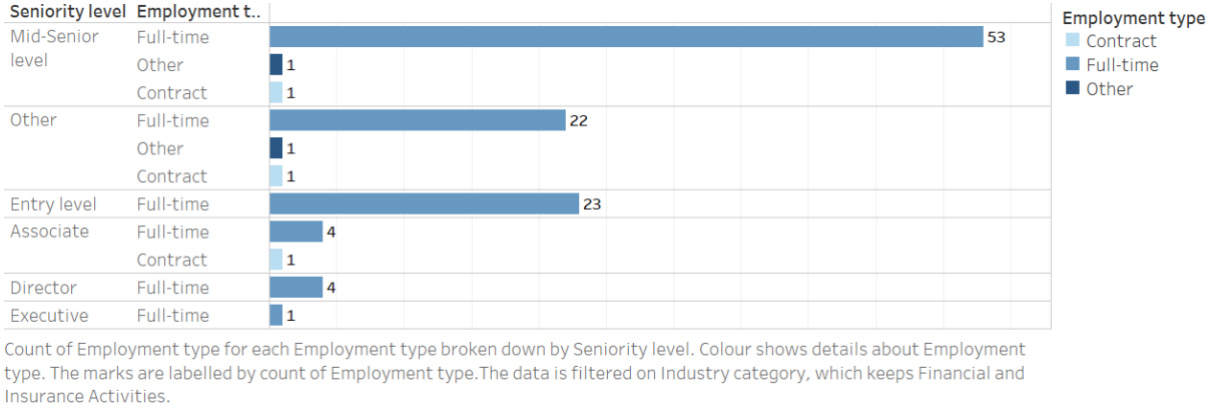


Figure 18: Financial and Insurance Activities Industry Seniority Level Vs Employment type (Source: Author)

E. Manufacturing

In the Manufacturing industry, the advertised job roles have a fairly equal mix of full-time and contract employment types at the mid-senior level (see figure 19). However, the associate role and other (which refers to the employment types that does not clearly fall into the classification) have only full-time roles. Data governance professionals in this industry at mid-senior level may get either contract or full-time roles, however other levels should focus on full time roles.

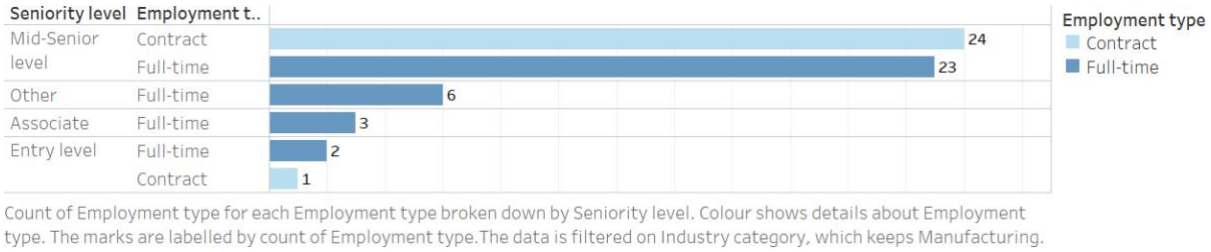


Figure 19: Manufacturing Industry Seniority Level Vs Employment Type (Source: Author)

4.1.8 Occurrence of the term "data governance" in Job title

92 advertised roles (9.74% of the advertised job titles) were found to have the term "data governance" in the job title (see figure 20). Although the number appears to be low, the availability of job roles with data governance in the job title may be indicative of the fact that there is an increasing awareness in organisations of the need for these data governance roles. Similarly, in line with literature on data governance roles and responsibilities, there are varying roles that do not include data governance in the title. Moreover, literature has shown that the jobs with data governance in the title are often related to the data governance office (Seiner 2019) .

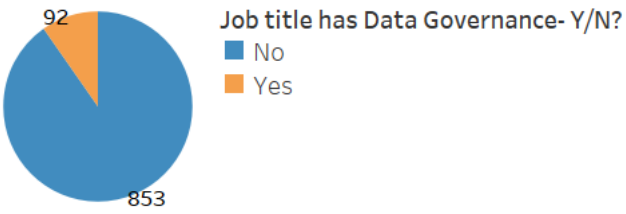


Figure 20: Occurrence of "data governance" in Job title (Source: Author)

4.1.9 Number of Occurrences of The Term "Data Governance" In Job Description

The term "data governance" was found in the job description 1957 times, however there is a wide variation in the emphasis placed on the term "data governance" in job descriptions across industries.

The occurrence of the term was observed to be higher in industries tied to data, finance, and professional services as seen from the figure. The number of occurrences of the term was found to be highest in the Information and Communication industry which may be considered as unsurprising because of the intrinsic relationship the industry has with data. Thus, these findings suggest that those seeking data governance roles should focus primarily on the information and communication, professional, scientific, and technical services, administrative and support services as well as the financial industries, as these appear to have a sizable demand for data governance professionals.

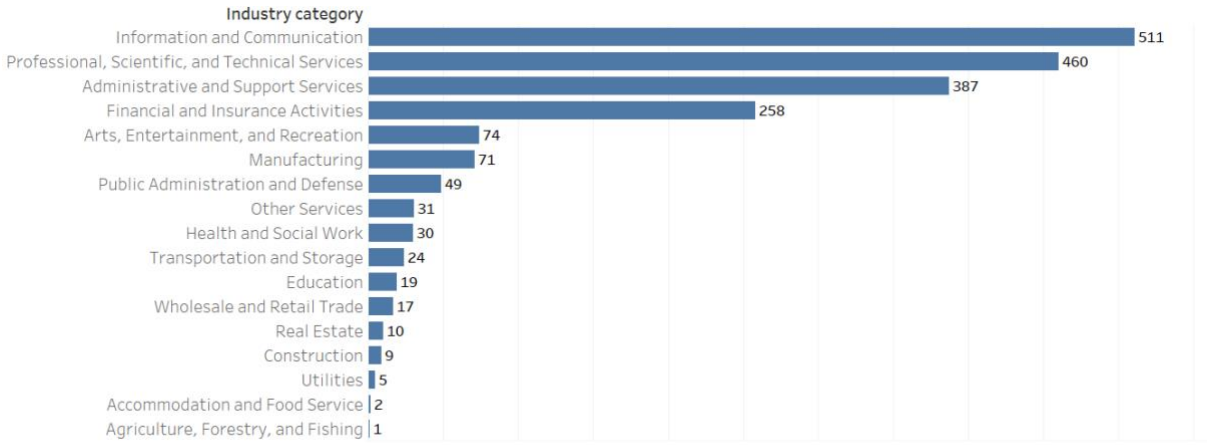


Figure 21: Occurrence of "data governance" in Job description (Source: Author)

It is interesting to note that twenty (20) advertised roles did not have data governance in the job description (see figure 22). Furthermore, the industry with the highest occurrence of this was the information and Communication industry which appears to have the highest number of advertised data governance roles. The key terms for these roles will be analysed and compared with that of the roles that have the occurrence of data governance in the job descriptions.

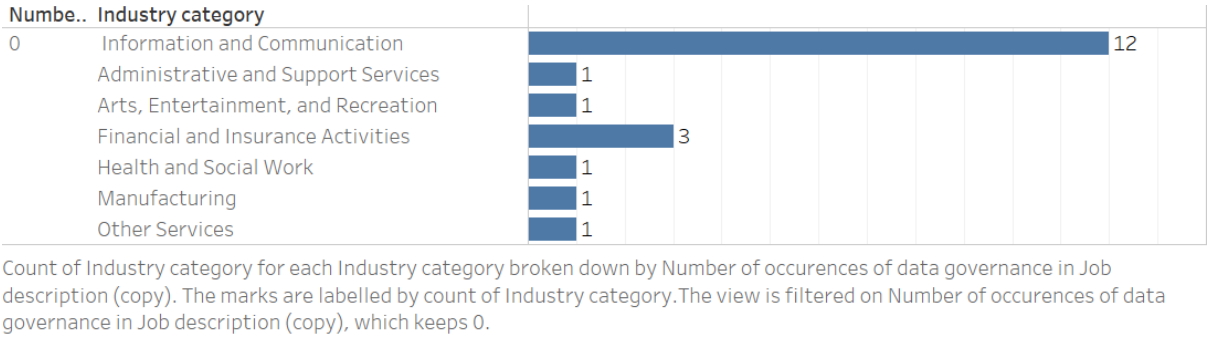


Figure 22: Breakdown of Job postings without "data governance" in job description by Industry (Source: Author)

4.2 Text Mining Analysis

Text mining produced various results related to the most frequently used terms. However, the most applicable results will be presented in this report and would include an analysis of the job titles as well as an analysis of the top 20 most frequent words across the top five industries used in the job description of the advertised roles.

4.2.1 Job Titles

The review of the job titles related to the data governance roles advertised show that there is a wide spectrum of job title nomenclature that reflects existing academic literature (See figure 23). Within these job titles, “data” appears as the most frequently occurring keyword highlighting the centrality of data, to data governance roles and initiatives. Additionally, other words such as “manager," "data engineer," "data architect," "analyst," "analytics," "Business Intelligence," "developer," "data quality," "privacy," and "strategy" frequently appear, aligning with the variety of roles and responsibilities described in academic literature.

Looking at the job titles described in the roles and responsibilities defined by Seiner (2019), terms like Manager, Lead, Senior Management, Head appears to be indicative of the individual with the responsibility for managing and implementing the data governance program within an organisation. These roles may be a part of the data governance council and sits in the executive or tactical level that has the responsibility of taking decisions on how data is handled in the organisation.

Also, since data governance implementation in organisations have multiple layers that are useful in addressing concerns faced at different levels of the organisation (Sebastian-Coleman 2018), terms such as business intelligence, analyst, data engineer, developer, project manager resonate more with the more operational level, likely aligning with the tasks of the working teams and specific governance-level functions.

Furthermore, these job titles align with the data governance domains identified in literature. Terms like Architect, data quality and security align with the roles applicable to the various domains identified by Abraham, Schneider and Vom Brocke (2019) and has also been identified as relevant roles to governing data in the Khatri and Brown framework (Khatri and Brown 2010). This overlap between academic research and job titles in practice serves as a guide for career-oriented decisions, pointing towards the skills and knowledge that are both academically sound and practically relevant.

This could be important to the aspiring data governance professional because an understanding of the nomenclature and semantics of data governance job titles may allow an individual to match their career goals and trajectories with organisational needs and theoretical frameworks.

Figure 23: Word Cloud of Job Title (Source: Author)

4.2.2 Top 20 terms in Job Description in Corpus

Figure 24 below shows the top 25 terms in the corpus of the job descriptions for roles that have the occurrence of data governance in the job descriptions. Words such as experience, business, work, team, management appear over 2000 times in the job postings and also in over 700 documents. These terms appear to be significant terms as they are used across most of the job roles advertised.

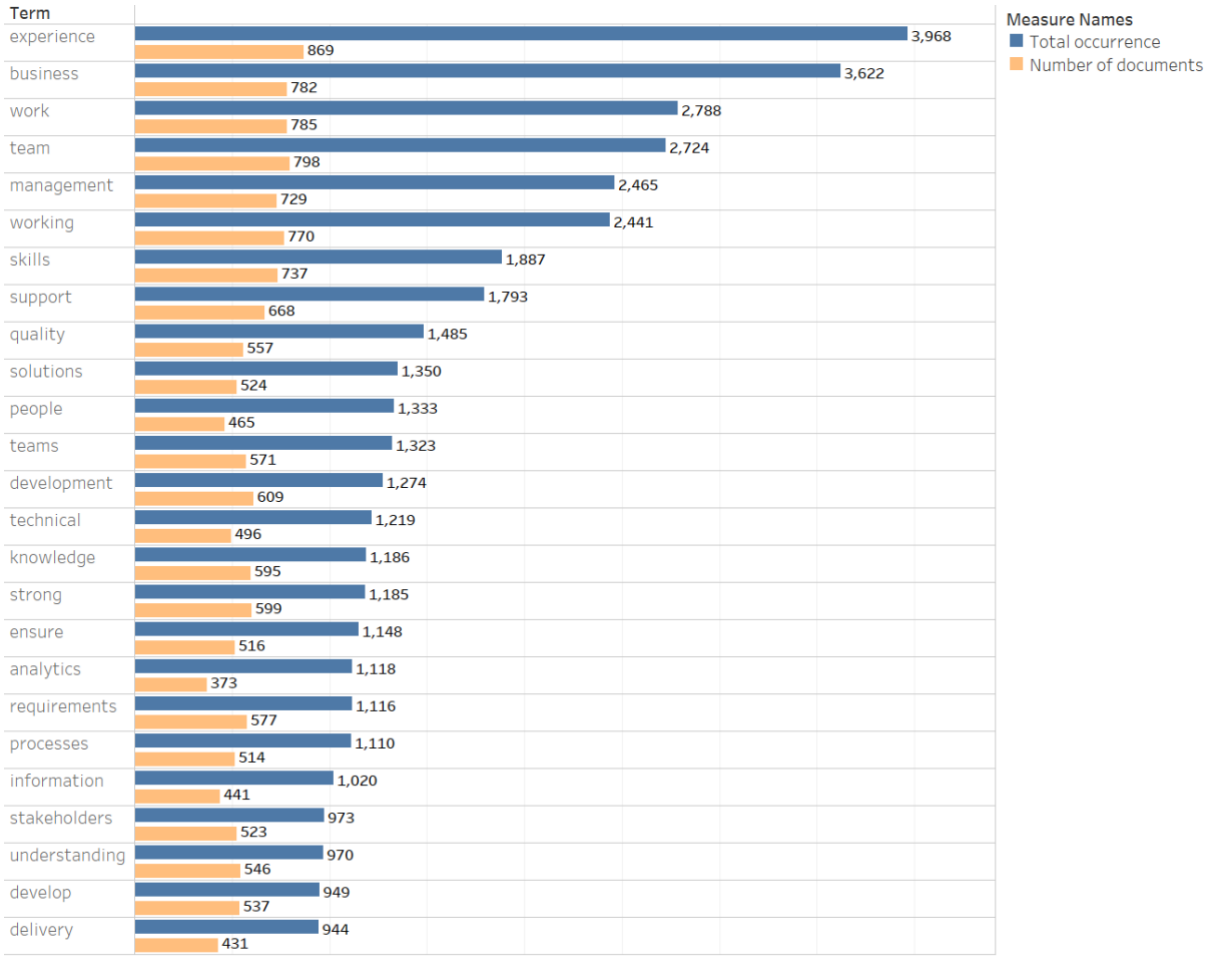


Figure 24: Top 25 Tokens (terms) in corpus of job description for roles that have the occurrence of the term "data governance" (Source: Author)

4.2.3 Top 20 terms in Job description across the Top 5 Industries

An analysis of the terms used in data governance job postings across various industries (see Table 5) may offer valuable insight into each sector's focus and priorities. Again, the roles in each industry may differ based on the terms used in the job postings; however, there are some terms that may be common to data governance job roles across all industries explored.

It was observed that the terms "work," "working," and "management" appear frequently across the top five industries, which highlights that they may be relevant to data governance job descriptions and roles.

Apart from the general terms found across all industries, some industries may have some unique terms. For instance, the information and communication industry used technology-related terms like “cloud”, “tools”, and “understanding” which could suggest a technical focus. On the other hand, the financial and insurance industry used terms such as “risk” and “financial” which may suggest an outlook on risk management, while the manufacturing industry used terms such as “systems” and “architecture” which may point to a focus on design and procedure. It was observed that there may have been a strong emphasis on "experience" in the Administrative and Support Services sector. This may mean that a wide range of skills are valued. Also, a significant number of references to "clients" and "firms" are found in the Professional, Scientific, and Technical Services industry, which may highlight the interaction with clients and possibly client data.

Term	Manufacturing	Administrative and support services	Financial and Insurance Activities	Professional, Scientific and Technical Services	Information and Communication
work	131	324	338	243	789
working	123	318	313	177	665
management	121	355	433	181	613
systems	87	-	-	-	-
skills	81	308	250	148	565
finance	74	-	-	-	-
support	74	181	263	143	474
teams	73	160	143	122	409
processes	71	-	141	-	-
architecture	65	-	-	-	-
information	65	-	143	-	-
product	64	-	-	-	-
requirements	60	202	170	-	318
development	58	165	-	115	377
make	56	-	-	-	-
solutions	56	203	182	118	502
lead	55	-	-	-	-
global	54	-	-	-	-
strong	53	197	157	-	372
design	51	-	-	-	-
experience	-	664	-	255	1210
business	-	547	526	312	984
team	-	404	315	209	715
analytics	-	245	-	-	301
quality	-	242	166	-	411
knowledge	-	185	154	102	384
technical	-	189	-	93	418

project	-	168	-	92	-
ensure	-	162	164	98	-
reporting	-	154	143	-	-
financial	-	-	162	-	-
people	-	-	156	140	345
risk	-	-	149	89	-
clients	-	-	-	170	-
firm	-	-	-	109	-
technology	-	-	-	107	-
tools	-	-	-	-	342
cloud	-	-	-	-	314
understanding	-	-	-	-	307

Table 5: Breakdown of Top 20 terms across Top 5 industries (Source: Author)

4.2.4 Data Governance Actions

Further analysis on the words was carried out based on Alhassan, Sammon and Daly's work that categorised data governance activities into three actions that show the "doing of data governance". These actions are named "define", "implement" and "monitor" actions. (Alhassan, Sammon and Daly 2016) Hence, each of the top terms identified in the job descriptions were analysed and categorised as one of the identified three actions.

A. Terms such as business, management, requirements, architecture, information, product, design, and understanding may be classified as "define" actions (see Table 6).

	Term	**Suggested Meaning**
1	Business	is a scope noun that may indicate that the business holds a fundamental role in defining data governance strategies i.e., data governance strategies and objectives are defined within the context of the overall business.
2	Management	may denote responsibility which could mean defining data governance policies as well as the responsibilities for the various aspects of data management in order to align with the overall business goals (Gupta and Cannon 2020b).
3	Requirements	may signify the definition of clear standards and criteria for handling the various aspects of data management within the organisation
4	Architecture	may relate to the design and definition of the enterprise data requirements, architectural policies, standards, and guidelines (Sebastian-Coleman 2018).
5	Information	within the context of data governance often refers to the data asset owned and managed by an organisation. Thus, the term information may suggest a focus on defining standards and guidelines for data governance frameworks and practices with respect to managing the data assets owned by the organisation.
6	Product	may refer to various artifacts, tools, and resources that may facilitate effective data management and governance practices such as data catalogues, metadata repositories, and data dictionaries. Hence, the term product may relate to defining these artifacts and tools with respect to the data management standards set within the organisation.
7	Design	may imply a focus on defining the data governance model in the organisation including the principles and guidelines for the frameworks and practices.

8	Experience	may denote the need for practical skills which may be relevant in defining data governance strategies.

Table 6: Suggested meanings for "define" action terms (Source: Author)

B. Terms such as work, working, systems, skills, support, teams, processes, solutions, lead, strong, development, make, technical, project and ensure may be categorised as “implement” actions (see Table 7).

	Term	**Suggested Meaning**
1	Work/ Working	may refer to a hands-on active involvement in the implementation of a data governance program or a data management process rather than a focus on the definition of policies and standards.
2	Systems	often refers to a set of parts or procedures working together to achieve a common objective and may imply the implementation of various data governance policies or practices.
3	Skills	are often considered relevant to any advertised role as the possession of the relevant skills is crucial to success in the role. In most cases, skills are used to accomplish tasks and achieve stated goals. Furthermore, it may imply that a data governance role requires a variety of skills.
4	Support	may typically denote assistance with the completion of setting tasks or actions. It may also refer to the day-to day implementation of actions that ensure that data governance strategies are adhered to.
5	Team/Teams	may denote teamwork particularly when paired with the term “work”. In a job description it may be used to denote the need for collaboration in order to achieve specific goals. In the context of the implement action, the term may be considered crucial because implementing policies, standards, processes are performed by teams. According to Seiner (2019), one of the teams described is a data governance working team that comes together to implement a specific action.
6	Processes	are often used to transform policies and standards into actionable tasks by defining the how. It could also refer to the process of putting data governance guidelines and standards into practise.
7	Solutions	may refer to the tools, software, strategies, or techniques used to implement data governance plans and may be used to denote tools, software or methodologies used in operationalising data governance.
8	Lead	may suggest the ability to take initiative towards the implementation of strategic objectives either from an organisation wide perspective or with respect to a data governance program.
9	Strong	is an adjective that may be used in a job description to emphasize the level of skill or experience that may be required for a role. As an implement action, it may be used to describe the effectiveness of an executed task.
10	Development	may generally be used to denote building which could also be described as an implementation of data governance strategies or policies.
11	Make	may denote the creation or contribution to something within the context of the job role. It may be further described as an implementation action that focuses on the active execution of actions that produces relevant results.
12	Technical	often relates to skills associated with the use of certain or specific technologies. It may point to the use of hands-on skills to implement data management strategies.

13	Project	may imply the need to plan and execute strategic tasks as a part of a broader organisational goal.
14	Ensure	may suggest a form of responsibility towards the completion of a successful implementation.
15	Technology/Tools / Cloud	may suggest the implementation of a technological tool in relation to addressing the data governance objectives within an organisation. Similarly, the implementation may refer to cloud technologies as well.
16	Understanding	often refers to having a grasp of concepts, tasks or technology which may imply the possession of an ability to implement certain concepts, tasks, or technology.

Table 7: Suggested meanings for "implement" action terms (Source: Author)

C. Terms such as analytics, quality, knowledge, financial, risk, clients, firm, reporting, people, and global may be categorised as “monitor” actions (see Table 8).

	Term	**Suggested Meaning**
1	Analytics	may be categorised as a monitor action as it is often used to track performance and effectiveness of implemented initiatives.
2	Quality	data quality is often monitored for adherence to the specified standards hence the term “quality” may be categorised as a monitor action.
3	Knowledge	may suggest a need for specialised expertise and may be described as a monitor action because possessing knowledge improves the ability to assess and measure the effectiveness of implemented strategic actions.
4	Financial	may often denote fiscal responsibility which may be relevant in monitoring organisational resources.
5	Risk	may imply a focus on risk assessment and mitigation which implies the occurrence of a monitoring activity in order to identify, assess and mitigate the risks involved in the implementation of a strategic action.
6	Clients	are crucial to every organisation. The term may suggest a need to ensure that data governance practices and regulations related to client’s data are closely monitored for adherence.
7	Firm/People/Global	may suggest a monitoring action of data related activities from an organisation wide or global perspective within the organisation.
8	Reporting	suggests that data management processes are reported on, to track and monitor performance.

Table 8: Suggested meanings for "monitor" action terms (Source: Author)

It is relevant to mention that some words due to their nature and based on usage within the specific job advertisement may be categorised as both define and implement actions or as both implement and monitor actions or as all three of the actions. However, the categorisation discussed above will be adopted for further analysis of the findings.

4.2.5 Data Governance actions across the Top 5 Industries

According to Alhassan, Sammon and Daly (2016), data governance actions in literature indicated that “implement” and “monitor” actions were higher in practice publications than in academic publications

while the define action is higher in academic publications than practice. An analysis of the top 20 terms across the top five industries suggests that the implement and monitor actions may be higher in practice as reflected from the job postings (see Figure 25).

The Information and Communications industry appears to use terms in the job postings that focus more on the implement action. This is closely followed by the monitor action with the define action being the lowest suggesting an almost equal focus on both implement and monitor actions in this industry. On the other hand, the Professional, scientific, and technical services industry exhibits a clear preference for the use of terms that fall into the implement action with the define and monitor actions being closer to each other suggesting a higher focus on the implement action and an almost equal focus on the define and monitor actions. The Administrative and Support Services Industry is similar to the Professional, Scientific and technical services industry however the terms categorised as define and monitor actions were the same while the Financial and Insurance activities industry also follows the trend of more terms focusing on the implement action. However, the monitor action is higher than the define action.

The Manufacturing industry has a slightly different pattern, in that while the occurrence of the terms categorised as implement action was higher, the define action was also higher than the monitor action. This may suggest a focus on definition of policies and programs as well as the subsequent implementation, but a lower focus may be on monitoring based on the advertised job roles.

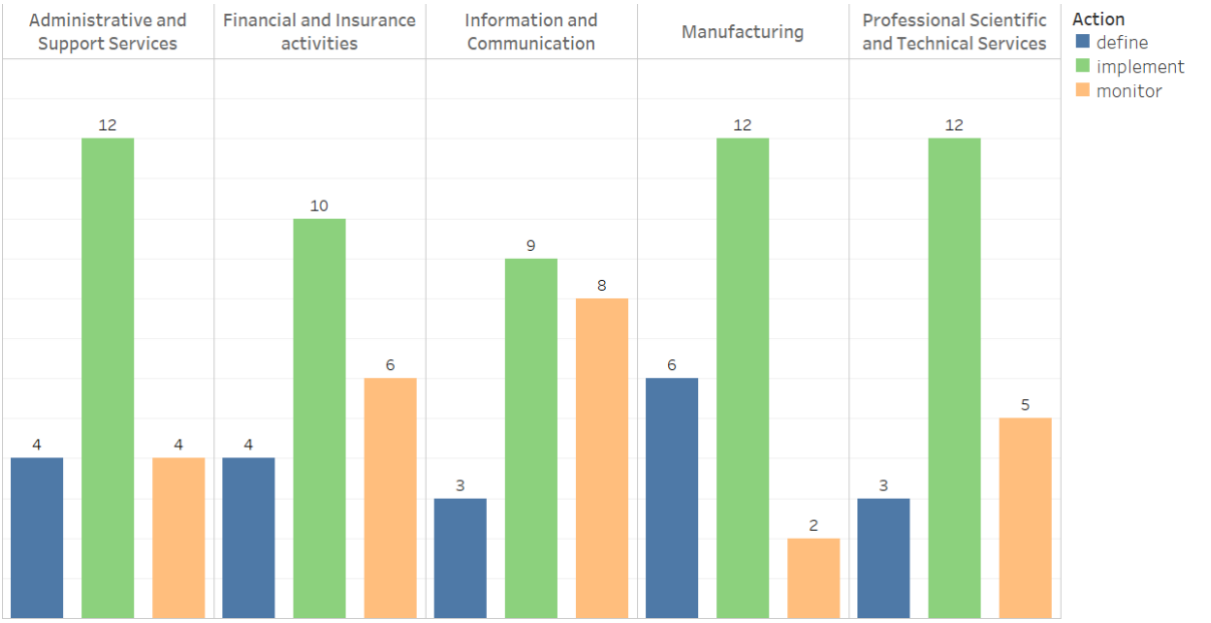


Figure 25: Data Governance Action across Top 5 Industries (Source: Author)

4.2.6 Job postings without “data governance” in job description

The top terms in the job adverts without the occurrence of data governance in job description was observed to be similar to the top terms in the job postings that had the occurrence of data governance in the job description (See Table 6). Terms such as team, experience, business, work, support, delivery, management were prevalent. However, it may be relevant to note that these terms occurred in over 50% of the relevant documents and may be considered as being relevant to the job postings.

Words such as environment which appeared in 13 out of the 20 documents was seen to be quite relevant and may refer to business, cultural, technical, or regulatory environments. This appears to be quite relevant to this category of jobs. Except for terms such as environment, there is no significant difference observed in the top terms used for jobs with the occurrence of data governance in the job description with those that do not have. This may suggest that data governance roles use similar terms regardless of whether data governance is mentioned or not.

term	In documents	Total
team	18	101
experience	17	86
business	16	85
work	19	74
support	19	67
delivery	14	65
management	13	56
environment	13	48
skills	16	45
working	18	43
teams	14	42
ensure	18	41
able	8	40
highly	13	38
opportunities	14	37
process	12	37
technical	13	37
clients	13	36
required	14	36
cloud	13	35

Table 9: Top 20 terms for job adverts without the occurrence of data governance in job description (Source: Author)

4.3 Discussion

The availability of data governance jobs advertised points to an increasing awareness of data governance in the United Kingdom which aligns with the UK data governance landscape in which the UK has set up various organisations, policies, and standards to ensure proper governance of data (The Royal Society 2020).

Research Question 1: What are the current trends and characteristics of the data governance job market in the United Kingdom?

The current trends and characteristics may be indicative of the overall job market analysis in the United Kingdom. England was observed to have the highest number of data governance job postings and Northern Ireland had the least (See Table 3). This aligns with the labour market trends (Office for National Statistics 2023) in which England had the highest employment rate, while Northern Ireland had the lowest employment rate. Similarly, London in the same report was identified as having a high employment rate in the UK, which aligns with the report that London was the Top city in which data governance roles were advertised (See figure 10). Employment types for these roles were mostly full-time roles and mid-senor level type roles. These characteristics may be indicative that some level of experience is required for data governance roles as well as that organisations may require some level of commitment, hence the full-time roles. This could be due to the multifaceted nature of the roles and the added complexity of formalising the processes around managing data.

Research Question 2: Which industries or sectors in the UK have the highest demand for data governance roles based on the number of job postings for data governance roles?

The study identified the following industries as having the top number of data governance job postings at the time of research (see figure 11).

i. Information and Communication,
ii. Administrative and support services,
iii. Professional, Scientific, and technical services
iv. Financial and insurance activities
v. Manufacturing

In comparison with other industries, these industries are diverse and might not have any features that distinguish them from the other industries. However, the common factors that may exist across these industries may be the use of data to guide decision making or treating data as an asset which makes

governing data critical to successfully running the business (Alhassan, Sammon and Daly 2019; Ladley 2019). Secondly, the degree of regulation across the industries may also be a factor influencing the demand for the data governance roles. Industries such as the financial industry may be considered to be highly regulated. However, based on the widespread knowledge of the UK data protection policy, it may also be argued that all industries in the country are extensively regulated. Furthermore, data governance as a field is suggested to have based its tenets on the principles of Information technology governance (Khatri and Brown 2010), hence it is perhaps unsurprising that the Information and communication industry has the highest number of advertised roles. Similarly, the access to and use of data in the industry may account for the higher number of job postings. However, the author did not find research that points to the difference in the demand for data governance professionals across various industries and this leaves room for further research.

Research Question 3: What is the number of job openings with "data governance" in the job title compared with the job openings without "data governance" in the job title?

The study revealed that 9.74% of the advertised roles at the time of the study had the term "data governance" in the job title. However, it may be suggested that the lack of the term "data governance" in the job title does not necessarily mean that the role is not related to data governance, as existing literature has shown that data governance roles and responsibilities are diverse and cuts across the various domains of data governance and data management (Abraham, Schneider and Vom Brocke 2019; Sebastian-Coleman 2018; Yulfitri 2016; Korhonen et al. 2014; Otto 2011b; Otto 2011a; Wende 2007). (Abraham, Schneider and Vom Brocke, 2019; Yulfitri 2016; Korhonen et al. 2014; Sebastian-Coleman 2018; Wende 2007; Otto 2011b; Otto 2011a). Furthermore, the presence of data governance in the job title may hint at roles which are related to the existence of a data governance office responsible for administering and facilitating the data governance program (Seiner 2019).

Research Question 4: What are the common/specific job titles and roles associated with data governance positions in the UK?

The findings showed that the job titles and roles associated with data governance positions in the UK are wide and varied though, they are similar to the roles stated in literature as discussed in Section 4.2.1. However, there is a noted absence of roles such as data steward which literature as shown to be a fundamental aspect of the operational level of the data governance roles and responsibilities (Yulfitri, 2016; Korhonen et al., 2014). This may relate to the argument by Seiner (2014) that data stewardship should not be a job title in itself or perhaps may be due to data stewards having multiple

responsibilities to the business, hence data governance may not be mentioned in job postings for the role but may be implied instead. On the other hand, the presence of job titles such as business analyst may also hint at data stewardship roles.

Research Question 5: What are the most frequent terms found in the job descriptions in data governance job postings?

The most frequent terms found in the job descriptions include terms such as experience, business, work, team and management which appear over 2000 times in the job postings and also in over 700 documents (see section 4.2.2 and 4.2.3). However, there is a notable absence of terms such as security, privacy, GDPR, policies and so on. These terms were found in the corpus but were not among the most frequently occurring words (See appendix III). The presence of these words in the corpus may suggest an awareness of data governance in organisations and a corresponding motivation to comply to regulations. However, the occurrence of the words, are low enough to suggest that practitioners may need to promote an awareness within organisations in the UK. In addition, it may be suggested that some responsibilities are implied and not explicitly stated within the content of job advertisements as literature has also found that there is some ambiguity around the definition of data governance roles and responsibilities (Karkošková 2023) hence organisations define them based on what their organisational goals and motivations may be.

This study defines data governance as the formal execution and enforcement of authority over the management of data and data related assets (Seiner 2014) which suggests that while data governance is about formalising the processes around governing data, there is an informal governance of data and data related assets within organisations referred to as informal data governance. This implies that while organisations may not have formal data governance programmes in place nor do they have formal data governance roles defined, there is a possibility that they however have some form of data governance implemented. On the other hand, from the non-invasive data governance approach (Seiner 2014), it is possible that organisations may choose to hire and grow the data governance roles from within the organisation, which may imply that the roles may exist, however may not be advertised.

Chapter Five: Conclusions and Recommendations

5.0 Introduction

The main aim of this research was to conduct a comprehensive analysis of the current state of the UK's employment market for data governance professionals, focusing on an examination of advertised data governance job roles using LinkedIn job postings as a source of information. This chapter provides a summary of the key findings in relation to the research objectives, literature review and constructs. Limitations of the study, recommendations as well as suggestions on possible future research was also included.

5.1 Summary of the Report and Conclusions

Data governance has become increasingly significant in recent times, with new job roles being created in organisations to manage data and data assets effectively. This study set out to gain an understanding of the current state of advertised data governance job roles in the United Kingdom in order to provide insights that may guide the career decisions of aspiring data governance professionals. The aim of the research and the research objectives was achieved as the data collected from LinkedIn was examined to reveal the country, cities, companies, employment type, seniority level and industries with the highest number of job postings for data governance roles. In addition, the job titles were examined alongside the most frequent terms found in the job descriptions. This was further compared across the top five industries with the most data governance job postings.

The findings of this study showed that the characteristics and trends of the UK employment market for data governance professionals appear to be in line with broader labour market trends as reported by the Office for National Statistics (2023). England, specifically London, had the highest number of advertised job postings with most jobs requiring mid-level seniority level and full-time roles. Secondly, the Information and communication industry was identified as having the highest number of data governance job postings which could be attributed to this being the “technology” industry and the reliance of this industry on data. Thirdly, a broad range of job titles was advertised with data governance included in some of the job titles, which could be indicative of an emerging awareness of formalising data governance as a distinctive function within organisations. In addition, it hints at the fact that the scope of data governance functions may extend beyond what is captured in the job title alone. Finally, the key terms in the job postings were identified and an analysis of these terms indicated that there may be more focus on the “implement” action in practice rather than the “define” and “monitor” actions.

Overall, the analysis of the data governance job advertisements suggests that organisations may predominantly be seeking experienced professionals with diverse skills who are able to offer consistency and continuity and also commit to organisational goals. In conclusion, the study has shown that there is a significant overlap between the academic literature for data governance and the job titles and roles that exist in practice. This overlap may serve as a practical guide for professionals aiming to match their skills and career objectives with both theoretical and real-world demands.

5.2 Limitations of Study

There are several limitations on this study. First data was collected in a single research session that had jobs posted for a period of thirty days since historical data of the job advertisements is not readily available on the platform. This limitation could be mitigated by periodical collection of job advertisements over a longer period of time which would provide a more comprehensive picture of the trends and patterns in the data governance roles in the UK. Furthermore, it would enable a better understanding of the demand for data governance roles in the UK.

Secondly, the data used for the study was job advertisements posted on LinkedIn. While LinkedIn has been acknowledged as being one of the leading websites for posting job advertisements, there are equally other similar sites that may be comparable to LinkedIn. Hence, there was a risk of bias introduced by using only one source of data. This could have been mitigated by collecting data on the job adverts from multiple job sites such as Indeed, Reed, Glassdoor, CVLibrary etc. In addition, since the source of data was only from online job postings, data governance job roles advertised internally and filled within organisations was excluded from the analysis. The bias may have been mitigated by conducting an interview or survey within organisations to examine the roles filled internally.

Thirdly, the text mining analysis focused on analysing term frequencies across the top industries identified by the number of job postings in the industry. There are other forms of natural language processing and text mining analysis available that may have provided further insights on the state of the data governance job market in the UK. However, this leaves room for further research using various methodologies to extract relevant insights.

Fourthly, some coding and interpretation of data carried out may be considered to be subjective and also dependent on the knowledge of the industry. This may be mitigated by an automated and less subjective process for analysing the terms.

5.3 Recommendations

Job advertisements are considered a useful source of information to HR professionals, higher institutions, and practitioners (both aspiring and those already in the field) as they can provide a mine of information regarding the state of the job market.

Based on the findings from the study, it could be recommended that aspiring data governance professionals in the United Kingdom focus their job search in England, and specifically London for optimal results. However, there may be opportunities in Edinburgh and Glasgow for professionals that would prefer Scotland. Secondly, since most roles are mid-senior level roles, it may be suggested that aspiring data governance professionals acquire experience in other parts of data management roles before going on to establishing a career in data governance. From an industry perspective, aspiring practitioners should focus on the Information and Communication industry although there may also be opportunities available across other industries. It is however recommended that experts implement measures to build awareness of formal data governance implementation across other industries.

There is a variation of job titles available to the aspiring data governance professional to enable them to develop the desired competence based on their career objectives. This is particularly relevant since terms like work and experience are common across the job descriptions.

5.4 Further research

This study opens up several opportunities for further research. First this study examined the current state of the advertised data governance roles in the UK employment market based on a single research session. Further studies can use data collected over a longer period of time with multiple data sources to establish trends and also quantify the demand for data governance professionals as research has shown that job advertisements are useful in quantifying demand for skills. This could be further extended by examining the data governance roles filled by organisations internally in order to build a clear picture of data governance roles and the key skills required by organisations for these roles.

Secondly, this study focused on an analysis of the top 20 terms generated from the job description. Further research could generate n-grams and analyse the key phrases generated in the job descriptions. Similarly, more complex text mining methodologies such as clustering and topic modelling may be applied to generate findings that will add value to human resource professionals, data governance professionals as well as educational institutions.

Acknowledgements

First, I would like to thank the Almighty God for the grace and inspiration in completing my research study.

Secondly, I also thank my supervisor Dr Sandra Cannon for her unending support and guidance throughout the process of writing the dissertation. She offered really helpful suggestions and advice that are really appreciated.

Thirdly, I would like to thank my family especially my husband and my mum, who provided a support system, believing in me and encouraging me even when I was tempted to give up.

Fourthly, I thank my friends some of whom have turned family, for being my sounding board for my thought process and the unending encouragement received.

Finally, I want to express my gratitude to the lecturers and module leaders who imparted me with knowledge during my time at RGU's Schools of Creative and Cultural Business, the study skills team particularly Dr Liam Waldron who taught me the brainstorming method that helped me to kick off my literature review after several attempts, the Library team, and my classmates/colleagues who were helpful and supportive at different times.

References

AARONSON, S., STRUETT, T. and ZABLE, A., 2022. *DataGovHub Paradigm for a Comprehensive Approach to Data Governance.* Medford, MA, USA: IDEA 2030 Council- Digital Planet.

ABRAHAM, R., SCHNEIDER, J. and VOM BROCKE, J., 2019. Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management,* 49, pp. 424-438.

ALHASSAN, I., SAMMON, D. and DALY, M., 2016. Data governance activities: an analysis of the literature. *Journal of Decision Systems,* 25, pp. 64-75.

ALHASSAN, I., SAMMON, D. and DALY, M., 2018. Data governance activities: A comparison between scientific and practice-oriented literature. *Journal of enterprise information management,* 31(2), pp. 300-316.

ALHASSAN, I., SAMMON, D. and DALY, M., 2019. Critical success factors for data governance: a theory building approach. *Information Systems Management,* 36(2), pp. 98-110.

AL-RUITHE, M., BENKHELIFA, E. and HAMEED, K., 2019. A systematic literature review of data governance and cloud data governance. *Personal and ubiquitous computing,* 23(5-6), pp. 839-859.

BEGG, C. and CAIRA, T., 2012. Exploring the SME Quandary: Data Governance in Practise in the Small to Medium-Sized Enterprise Sector. *Electronic Journal of Information Systems Evaluation,* 15(1), pp. pp3-13.

BRANKA, J., 2018. *Jobs and skills: An online match.* [online]. Thessaloniki, Greece: European Centre for the Development of Vocational Training. Available from: https://www.cedefop.europa.eu/en/blog-articles/jobs-and-skills-online-match [Accessed Jun 27, 2023].

CHEONG, L.K. and CHANG, V., 2007. The need for data governance: a case study. *ACIS 2007 proceedings,* 100, pp. 999-1008.

CHOUDHURY, D., 2014. Data Governance Good Practices and the role of Chief Information Officer. *International Journal of Academic Research,* 1(3), pp. 176-183.

CRESWELL, J.W. and CRESWELL, J.D., 2023. *Research design : Qualitative, quantitative, and mixed methods approaches.* Sixth ed. London: SAGE.

DAVIES, N., 2021. *5 Common Data Governance Challenges (and How to Overcome Them).* [online]. Studio City, CA: Dataversity. Available from: https://www.dataversity.net/5-common-data-governance-challenges-and-how-to-overcome-them/ [Accessed Jun 22, 2023].

EUROPEAN COMMISSION, 2023. *European data strategy.* [online]. Luxembourg: European Union. Available from: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en [Accessed Jun 29, 2023].

FENNELL, A., 2021. *20 Best job sites in the UK [Find your perfect job today].* [online]. London: StandOut CV. Available from: https://standout-cv.com/pages/best-job-sites-uk [Accessed Aug 2, 2023].

FLECKENSTEIN, M. and FELLOWS, L., 2018. *Modern data strategy.* Cham, Switzerland: Springer International Publishing AG.

FLYVERBOM, M., DEIBERT, R. and MATTEN, D., 2019. The governance of digital technology, big data, and the internet: New roles and responsibilities for business. *Business & Society,* 58(1), pp. 3-19.

GOV.UK, 2020. *National Data Strategy.* [online]. London: UK Government. Available from: https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy [Accessed May 31, 2023].

GOV.UK, 2022. *UK Digital Strategy.* [online]. London: UK Government. Available from: https://www.gov.uk/government/publications/uks-digital-strategy/uk-digital-strategy [Accessed Jun 13, 2023].

GOV.UK, 2023a. *British Businesses to Save Billions Under New UK Version of GDPR.* [online]. London: UK Government. Available from: https://www.gov.uk/government/news/british-businesses-to-save-billions-under-new-uk-version-of-gdpr [Accessed Jun 2, 2023].

GOV.UK, 2023b. *Data governance panel formed to improve use of court and tribunals data.* [online]. London: UK Government. Available from: https://www.gov.uk/government/news/data-governance-panel-formed-to-improve-use-of-court-and-tribunals-data [Accessed Jun 2, 2023].

GRBICH, C., 2013. *Qualitative data analysis : An introduction.* London: SAGE Publications Ltd.

GREGORY, A., 2011. Data governance—Protecting and unleashing the value of your customer data assets: Stage 1: Understanding data governance and your current data management capability. *Journal of Direct, Data and Digital Marketing Practice,* 12, pp. 230-248.

GREGORY, A. and BENTALL, L., 2012. Data governance—Protecting and managing the value of your customer data assets: Stage 3: Identifying and controlling the risk in using third-party processors. *Journal of Direct, Data and Digital Marketing Practice,* 13, pp. 335-344.

GUPTA, U.G. and CANNON, S., 2020a. *A practitioner's guide to data governance : A case-based approach.* Bingley, England: Emerald Publishing.

GUPTA, U.G. and CANNON, S., 2020b. A Review of Data Governance Definitions and Emerging Perspectives. *International Journal of Data Analytics,* (1, 2), pp. 30-47.

HANEEM, F. et al., 2019. Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management,* 45, pp. 25-43.

HARNHAM, 2023. *Data Governance Jobs.* [online]. Wimbledon: Harnham. Available from: https://www.harnham.com/data-technology-overview/data-governance-jobs/ [Accessed Aug 21, 2023].

HARVARD BUSINESS REVIEW, 2021. *Reengineering the Recruitment Process.* [online]. Brighton, MA: HBR Press. Available from: https://hbr.org/2021/03/reengineering-the-recruitment-process [Accessed Jun 27, 2023].

HOOI-TEN WONG, D., MAAROP, N. and SAMY, G.N., 2020. Data governance and data stewardship: A success procedure. *2020 8th international conference on information technology and multimedia (ICIMU).* 2020. IEEE, pp. 54-61.

HOPWOOD, P., 2008. Data governance: One size does not fit all. *Information Management,* 18(6), pp. 16.

IBM CORPORATION, 2007. *The IBM data governance blueprint: Leveraging best practices and proven technologies.* Armonk, NY: IBM Corporation.

IGNATOW, G. and MIHALCEA, R., 2018. *An introduction to text mining: Research design, data collection, and analysis.* Los Angeles, CA: SAGE Publications, Inc.

JAGGIA, S. et al., 2020. Applying the CRISP-DM Framework for Teaching Business Analytics. *Decision sciences journal of innovative education,* 18(4), pp. 612-634.

JANG, K. and KIM, W., 2021. Development of data governance components using DEMATEL and content analysis. *The Journal of supercomputing,* 77(4), pp. 3695-3709.

JOHNSON, R.B., ONWUEGBUZIE, A.J. and TURNER, L.A., 2007. Toward a definition of mixed methods research. *Journal of mixed methods research,* 1(2), pp. 112-133.

KARKOŠKOVÁ, S., 2023. Data Governance Model To Enhance Data Quality In Financial Institutions. *Information Systems Management,* 40(1), pp. 90-110.

KHATRI, V. and BROWN, C.V., 2010. Designing data governance. *Communications of the ACM,* 53(1), pp. 148-152.

KNIGHT, M., 2021. *What Is Data Governance?* [online]. Studio City, CA: Dataversity. Available from: https://www.dataversity.net/what-is-data-governance/ [Accessed Jun 5, 2023].

KOLTAY, T., 2016. Data governance, data literacy and the management of data quality. *IFLA journal,* 42(4), pp. 303-312.

KORHONEN, J.J. et al., 2014. Designing data governance structure: an organizational perspective. *GSTF Journal on Computing,* 2(4), pp. 11-17.

KOTU, V. and DESHPANDE, B., 2014. *Predictive analytics and data mining: Concepts and practice with RapidMiner.* San Francisco: Elsevier Science.

KURGAN, L.A. and MUSILEK, P., 2006. A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review,* 21(1), pp. 1-24.

LADLEY, J., 2019. *Data governance how to design, deploy and sustain an effective data governance program.* 2nd ed. Cambridge, MA: Academic Press.

LEONE, M. and WALKER, K., 2022. *ESG Research Insights Paper: 2022 State of Data Governance and Empowerment.* Newton, MA: ESG.

LILLIE, T. and EYBERS, S., 2018. Identifying the constructs and agile capabilities of data governance and data management: A review of the literature. In: K. KRAUSS, M. TURPIN and F. NAUDE, eds. *Idia 2018. Proceedings of the Locally Relevant ICT Research* 23-24 August 2018. Cham: Springer International Publishing, pp. 313-326.

LIS, D. and OTTO, B., 2020. Data governance in data ecosystems - insights from organizations. *Strategic and competitive use of IT. Proceedings of the AMCIS 2020 Proceedings* 10 August 2020. Salt Lake City, UT: AMCIS, pp. 1-10.

MACAPINLAC, T., 2019. The Legality of Web Scraping: A Proposal. *Federal Communications Law Journal,* 71(3), pp. 399-0_8.

MAHANTI, R., 2021. *Data governance and data management : Contextualizing data governance drivers, technologies, and tools.* Singapore: Springer.

MARTINEZ-PLUMED, F. et al., 2021. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering,* 33(8), pp. 3048-3061.

MCGILVRAY, D., 2021. *Executing data quality projects: Ten steps to quality data and trusted information (TM).* Cambridge, MA: Academic Press.

MEDEIROS, M.M., MAÇADA, A.C. and HOPPEN, N., 2021. The role of big data stewardship and analytics as enablers of corporate performance management. *Revista de Administração Mackenzie,* 22, pp. eRAMD210063.

HANDEL, MICHAEL J., 2012. *Trends in Job Skill Demands in OECD Countries. OECD Social, Employment and Migration Working Papers.* Paris, France: OECD

MORGAN, D.L., 2017. *Integrating qualitative and quantitative methods : A pragmatic approach.* London: SAGE Publications Ltd.

OFFICE FOR NATIONAL STATISTICS, 2023. *Labour market in the regions of the UK - Office for National Statistics.* [online]. Newport: Office for National Statistics. Available from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/regionallabourmarket/august2023 [Accessed Sep 3, 2023].

OTTO, B., 2011a. Data Governance. *Business & Information Systems Engineering,* 3, pp. 241-244.

OTTO, B., 2011. A morphology of the organisation of data governance. *European conference on information systems. Proceedings of the ECIS* 10 June 2011. Helsinki, Finland: Association for Information Systems Electronic Library, pp. 272

PEJIC-BACH, M. et al., 2020. Text mining of industry 4.0 job advertisements. *International Journal of Information Management,* 50, pp. 416-431.

PLOTKIN, D., 2020. *Data stewardship: An actionable guide to effective data management and data governance*. Cambridge, *MA:* Academic press.

REDMAN, T.C., 2013. *Data's Credibility Problem.* [online]. Brighton, MA: HBR Press. Available from: https://hbr.org/2013/12/datas-credibility-problem [Accessed Jul 3, 2023].

RIGGINS, F.J. and KLAMM, B.K., 2017. Data governance case at KrauseMcMahon LLP in an era of self-service BI and Big Data. *Journal of Accounting Education,* 38, pp. 23-36.

SALTZ, J.S., 2021. Crisp-dm for data science: Strengths, weaknesses and potential next steps. *2021 IEEE international conference on big data (big data).* 2021. IEEE, pp. 2337-2344.

SCHEDLBAUER, J., RAPTIS, G. and LUDWIG, B., 2021. Medical informatics labor market analysis using web crawling, web scraping, and text mining. *International journal of medical informatics,* 150, pp. 104453.

SCHNEIDER, S. and KOKSHAGINA, O., 2021. Digital transformation: What we have learned (thus far) and what is next. *Creativity and innovation management,* 30(2), pp. 384-411.

SEBASTIAN-COLEMAN, L., 2018. *Navigating the labyrinth: An executive guide to data management.* Basking Ridge, NJ: Technics Publications.

SEINER, R.S., 2014. *Non-invasive data governance: The path of least resistance and greatest success.* First ed. Basking Ridge, NJ: Technics Publications.

SEINER, R.S., 2019. *Data Governance Roles and Responsibilities.* [online]. Pittsburgh, PA: Robert S. Seiner. Available from: https://tdan.com/data-governance-roles-and-responsibilities/24774 [Accessed Jul 20, 2023].

SHMUELI, G. et al., 2023. *Machine learning for business analytics : Concepts, techniques and applications in RapidMiner.* Hoboken, NJ: Wiley.

TALEND, 2023. *What is Data Governance (and Do I Need It)?* [online]. San Mateo, CA: Talend. Available from: https://www.talend.com/uk/resources/what-is-data-governance/ [Accessed Jul 20, 2023].

TALIB, R. et al., 2016. Text mining: techniques, applications and issues. *International Journal of Advanced Computer Science and Applications,* 7(11), pp. 414-418.

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(3), pp. 141-178.

THAKUR, K. and KUMAR, V., 2022. Application of text mining techniques on scholarly research articles: Methods and tools. *New Review of Academic Librarianship,* 28(3), pp. 279-302.

THE DATA GOVERNANCE INSTITUTE, 2023a. *Definitions of Data Governance.* [online]. Washington, DC: The Data Governance Institute. Available from: https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/ [Accessed Jul 13, 2023].

THE DATA GOVERNANCE INSTITUTE, 2023b. *DGI Data Governance Framework.* [online]. Washington, DC: The Data Governance Institute. Available from: https://datagovernance.com/the-dgi-data-governance-framework/ [Accessed Jul 12, 2023].

THE ROYAL SOCIETY, 2020. *The UK data governance landscape.* London, United Kingdom: The Royal Society.

UPADHYAY, H. et al., 2023. COVID-19, jobs and skills-Exploratory analysis of the job postings in the US and UK healthcare job market. *PLoS One,* 18(1), pp. e0278237.

UPWORK, 2022. *The Best Web Scraping Tools in 2023 | Upwork.* [online]. San Francisco, CA: Upwork. Available from: https://www.upwork.com/en-gb/resources/best-web-scraping-tools [Accessed Aug 2, 2023].

VAN VELDHOVEN, Z. and VANTHIENEN, J., 2022. Digital transformation as an interaction-driven perspective between business, society, and technology. *Electronic Markets,* 32(2), pp. 629-644.

VIAL, G., 2023. Data governance and digital innovation: A translational account of practitioner issues for IS research. *Information and organization,* 33(1), pp. 100450.

VOIGT, P. and VON DEM BUSSCHE, A., 2017. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing,* 10(3152676), pp. 10-5555.

WALSH, M.J., MCAVOY, J. and SAMMON, D., 2022. Grounding data governance motivations: a review of the literature. *Journal of decision systems,* 31, pp. 282-298.

WARD-PROUD, L., 2014. Back to basics: End of the job advert. *City A.M*, 04 June pp. 22.

WEBER, K., OTTO, B. and ÖSTERLE, H., 2009. One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality,* 1(1), pp. 1-27.

WENDE, K., 2007. A model for data governance–Organising accountabilities for data quality management*. Proceedings of the 18th Australasian Conference on Information Systems* 06 December 2007. Toowoomba, Australia: ACIS, pp. 417-425.

WILLIAMSON, K. and JOHANSON, G., 2017. *Research methods : Information, systems and contexts.* Oxford: Chandos Publishing.

YULFITRI, A., 2016. Modeling operational model of data governance in government: Case study: Government agency X in jakarta. *2016 international conference on information technology systems and innovation.* 2016. IEEE, pp. 1-5.

ZAKI, M., 2019. Digital transformation: harnessing digital technologies for the next generation of services. *Journal of Services Marketing,* 33(4), pp. 429-435.

ZHANG, Q., SUN, X. and ZHANG, M., 2022. Data matters: A strategic action framework for data governance. *Information & management,* 59(4), pp. 103642.

ZHANG, Y., CHEN, M. and LIU, L., 2015. A review on text mining. *2015 6th IEEE international conference on software engineering and service science (ICSESS).* 2015. IEEE, pp. 681-685.