REPUBLIC OF TURKEY

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technology

THE IMPACT OF BIG DATA FEATURES ON DATA INTEROPERABILITY CONCEPT

Sana Ibrahim SHATWAN

Master's Thesis

Supervisor

Asst.Prof.Dr Oğuz ATA

Istanbul, 2022

THE IMPACT OF BIG DATA FEATURES ON DATA INTEROPERABILITY CONCEPT

Sana Ibrahim SHATWAN

Information Technology

Master's Thesis

ALTINBAŞ UNIVERSITY

Istanbul, 2022

The thesis titled THE IMPACT OF BIG DATA FEATURES ON DATA INTEROPERABILITY CONCEPT prepared by SANA SHATWAN submitted on 09 /08 /2022 and has been **accepted unanimously** for the degree of Master of Science in Infromation Technology.

_____

Asst.Prof.Dr Oğuz ATA

Supervisor

Thesis Defense Committee Members:

Asst.Prof.Dr Oğuz ATA	Faculty of Engineering and Architecture, Altınbaş Univesity	_____
Asst.Prof.Dr. Doğu Çağdaş ATİLLA	Faculty of Engineering and Architecture, Altınbaş University	_____
Asst.Prof.Dr. Aytuğ BOYACI	Faculty of Air Force Academy, National Defense University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to the Institute of Graduate Studies: __/__/___

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

SANA SHATWAN

DEDICATION

I extend my sincere thanks and gratitude my father and mother for their unconditional support from the beginning of my journey in this life until now. I thank god for giving me a wonderful soulmate who always supports me, my dear husband.

ABSTRACT

THE IMPACT OF BIG DATA FEATURES ON DATA INTEROPERABILITY CONCEPT

Shatwan, Sana,

M.Sc, Information System, Altınbaş University,

Supervisor: Asst.Prof.Dr Oğuz Ata

Date: August 2022

Pages: 65

Big data has become an exciting term in the last decades, and it conducts a significant rate of data with different features of velocity, variety, value, volume, and veracity. Those features describe big data identity. Think about the data generated by social media websites and apps, such as Facebook, Twitter, and WhatsApp…...Etc. This data is crucial because of the amount, variety, and growth acceleration, but it is exhausting to manage. Big data interoperability indicates the capability to interchange and shares the data, information, and knowledge between organizations and devices that knowledge and information were extracted from different data sources. This complex mutual process grows an organization's performance. According to the previous, big data and its five features are essential because of its usability and analytics. Organizations need to extract the expected benefit from preserving and retrieving data. We develop a data interoperability concept, which is improved as data exchangeability can be enriched. In this study, we will discuss the impact of big data features value, volume, variety, velocity, and veracity on the interoperability of

the data and data exchangeability feasibility and how these features affect the prediction of future decisions in the institution, as well as the interaction and exchange of data between the organization and other institutions.

Keywords: Big Data, Big Data Interoperability, Big Data Features.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

ALCPO	:	Arrageeb Libyan Consumer Protection Organization.
OIMM	:	The organizational interoperability maturity model.
MMEI	:	The maturity model for enterprise interoperability.
XML	:	Extensible markup language.
DBMS	:	Database management system.
IOT	:	Internet of things.
ETL process	:	Extract transfer load.
EII	:	Enterprise's information integration.
SQL	:	Sequence query language
ICT	:	Information communication technology infrastructure
VIF	:	Variance inflation factor

LIST OF SYMBOLS

InteropC	:	The interoperability Cost.
E.C.	:	Expected cost.
R.C.	:	Real cost.
InteropT	:	The interoperability Time of the exchange.
R.T.	:	The real-time exchange.
E.T.	:	The expected time of the exchange.
InteropQ	:	Quality of the exchange.
A.E.	:	Amount of succeeded exchange.
T.E.	:	The expected time of the exchange.
C.E.	:	Conformity of the exchange.
A.C.	:	Amount of information that is conforming.
AR	:	Amount of information received.

1. INTRODUCTION

Big data could define according to different aspects of information, technology and impact. For instance, big data is defined as a massive amount of data, and diverse information assets that need profitable, creative models of information manipulation to increase clarity and decision-making [1]. Another official qualifier classified big data according to four categories [2].

a) Attributes of data: where five dimensions are indicated.
b) Technological needs: a structure that can develop adequate storage, exploit and analyze data.
c) Thresholds: exceeding the processing capacity.
d) Social impact: big data is a civilized, technological, and universal phenomenon.

Throughout data attributes, four features define Big Data: volume, velocity, variety, and value. Any height in volume, velocity, and variety of information assets demands cost-effective, creative forms of information processing for enhanced insight and decision-making [2]. Then, with technology concerns, Extensive data sets, primarily in the characteristics of volume, velocity, and/or variety, demand a scalable architecture for efficient storage, manipulation, and analysis. [2] also, another definition thresholds the data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis, and visualization technologies [2]. For social impact, big data is a phenomenon that causes three major transformations in how we evaluate data, transforming how we comprehend and manage society.:

a) More data.
b) Incomplete data.
c) Correlation overtakes causality [2].

So, as stated before, big data is not a traditional set of data, and it has different features (velocity, variety, value, volume, veracity). These features give different properties to the data. So, through these characteristics, the data will be described as big data. Therefore, to earn big data benefits, must be efficient and flexible to deal with heterogeneous systems. Exchanging data needs to conduct another essential concept, big data interoperability.

Big data interoperability has an impact on the whole organization's performance. This big data interoperability is tough and complex to be implemented with heterogeneity. Also, many barriers have appeared as follows:

a) Organizational barrier (refers to adjusting an organization's business process, responsibilities, and personal data protection).
b) Legal barrier (different organizations mean diverse operating platforms with different strategies and policies).
c) Technical barrier (concerns with different technical aspects of applications and I.T. infrastructure like (Semantic, syntactic, data inconsistency, data quality).

So, organizations must improve their data interoperability by taking into account different features of big data.

1.1 PROBLEM STATEMENT

Arrageeb Libyan Consumer Protection Organization (ALCPO) has many tasks, some of them as follows:

a) Educate Consumers and improve their awareness.
b) Collecting data on different products.
c) Legal assistance, as advice consumers legally and filing a lawsuit.
d) Encourage the consumers to bear an appropriate reaction towards penetrations and breaches.
e) Publishing Periodical instructions, prices, and news.
f) Consumer protection plan.

So, the study will discuss the characteristics of big data in the context of the following:

a) Value: The generated Data has to be valuable, reliable, and trustworthy data.
b) Volume: A massive amount of data has been generated.
c) Variety: Indicates the diversity of data (structure, unstructured, and semi-structured).
d) Velocity: Indicates the speed of data creation and exchange.
e) Veracity: It is about the biases, noise, and abnormality in data.

According to the previous, the contribution will be as follows:

a) Achieve the characteristics of big data with detailed explanations.
b) Measure the impact of the previous characteristics on the data interoperability of the organization.
c) Explain the results and how the interoperability of big data will be impacted by big data characteristics.
d) The results of the study will be illustrated.

1.2 STUDY HYPOTHESES

a) There is a significant effect of big data features on interoperability.

i. There is a considerable effect of the data value on the interoperability of the data.
ii. There is a considerable effect of the data volume on the interoperability of the data.
iii. There is a considerable effect of the data variety on the interoperability of the data.
iv. There is a considerable effect of the data velocity on the interoperability of the data.
v. There is a considerable effect of the data Veracity on the interoperability of the data.

b) There is a considerable positive correlation between big data features and data interoperability.

i. There is a considerable positive correlation between data value and interoperability.
ii. There is a considerable positive correlation between data volume and interoperability.
iii. There is a considerable positive correlation between data variety and interoperability.
iv. There is a considerable positive correlation between data velocity and interoperability.
v. There is a considerable positive correlation between data Veracity and Interoperability.

1.3 STUDY VARIABLES

a) The independent variable: is the effect of big data features represented in five dimensions (The value, The volume, The variety, The velocity, The Veracity).

b) The dependent variable: is the effectiveness of data interoperability, considering the following (data interoperability potentiality, data interoperability compatibility, data interoperability performance).

1.4 OBJECTIVE OF THE STUDY

A collected data from daily events is considered an important resource to determine many criteria related to ALPCO and how they will affect interoperability. When that data is compatible with other event evidence, both will help construct a collective view of a specific case. In contrast, the accumulated data becomes big data in a short time. So, it is necessary to preserve this data, which will be beneficial to analyze an event input and then propose a decision. Interoperability is crucial to improving the whole organization's operation process, business function, and decision-making, taking a stance on big data analytics.

This study aims to measure the impact of big data characteristics on the whole organization's data interoperability.

a) Categories and preserve those massive emitted data from ALPCO.
b) Analyze the collected data and evidence (reports, personal investigations, and other evidence and documents).
c) Measure the effect of big data characteristics on data interoperability.
d) Describe how the data interoperability will be impacted by big data characteristics.
e) Prepare conclusion report.

The study outcomes provide a better understanding and explanation of how big data five features will affect the data interoperability of the organization.

1.5 STUDY MOTIVATION

Nowadays, it is clear that data interoperability is important to deal with other necessary procedures like data integrity—for instance, time and cost of information exchange, etc. So, improving those procedures is essential to achieving data interoperability.

Hereby talking about big data interoperability, massive data and lots of documents will be generated from organizations. So, it is hard to process and give final results and reports and

decide in the same situation. Here, the question is who will manage all of these data and who can get a conclusion and meaningful results that we can use to achieve our benefit goal. This study represents the main items that can describe big data features and their impacts on data interoperability.

1.6 SUBJECT AND SCOPE

Big data features describe the different aspects of big data, distinguishing big data from traditional data. Also, data interoperability has an essential rule to exchange data and services between different heterogeneous systems. So, the value, volume, variety, velocity, Veracity, and their impact on the exchanging features of big data has to be discussed as follows:

a) Collecting data (Questionnaire, interviews, documents, observations, etc.)
b) Preprocessing data (make data clear and ready for the next stage).
c) Analyze the data.
d) Figure the result and conclusion.

1.7 LITERATURE REVIEW/MARKET SURVEY

Data interoperability and its barriers become narratively conducted because of their impact on information systems and data reciprocity. This could impact all organizations' performance. According to the previous, the interest goes to both majors, academic and practical fields. So, many different techniques and approaches become essential for future research in data interoperability, big data features, and decision making:

a) Interoperability measurements (Potentiality, compatibility, performance).
b) Framework, generic API, federated, and model-driven architecture-based approach.
c) The organizational interoperability maturity model (OIMM).
d) The maturity model for enterprise interoperability (MMEI).
e) Rational and random choice decision-making tools.
f) Decision matrix and weighted decision matrix.

1.7.1 Literature Review:

i. Maryam Ghasemaghaei, Goran calic (2020). Assessing the impact of big data on firm innovation performance: Big Data is not always better data. Journal of Business Research 108(2020) 147-162 [3].

Maryam and Goran discuss the impact of three big data features (volume, velocity, and variety) on innovation performance (efficiency and efficacy). The results show that instead of focusing on handling big data features as a holistic variable, it is important to negotiate the conceptual and operational differences among those features. The study implies that velocity and variety positively affect a firm's innovation performance, and data velocity has a more substantial role than the others. On the contrary, data volume does not play that significant role. And it is crucial to negotiate big data features as separate concerns, which will enhance overall innovation performance. Also, they conducted that big data is not always the best data.

ii. Maryam Ghasemaghaei (2019). Understanding the impact of big data on firm performance: the necessity of conceptually differentiating among big data characteristics. International Journal of Information Management [6].

This study reveals the effect of essential big data characteristics (Volume, Velocity, and Variety) on firm performance and the mediating role of data veracity and value to stand that relation. The findings present that variety improves data value, whereas the volume and velocity do not impact. Also, veracity has a negative effect on data volume, but variety and velocity are affected positively. This means data value has entirely arbitrated the impact of veracity on firm performance. These results support firms' managers to improve a whole firm performance and competitive advantage.

iii. Nir Kshetri (2014). Big data's impact on privacy, security, and consumer welfare. International Journal of Information Management. The University of North Carolina [4].

The study investigates the relation between big data characteristics, privacy, security, and consumer welfare. This relation is examined from different sides (data collection, storing, sharing, and accessibility). Using big data and specific features and linking them with gathering, storing, manipulating, and being accessible by another partner may lead to being stuck with many privacy and security issues. The companies and consumers will be concerned about potential offences and misuse of their information. Also, the consumers will not be comfortable as companies know and use more than they intentionally provide. Another concern that will negatively affect consumer welfare is unskilled and technologically inexperienced consumers. Those can cause many defects like (awareness of multiple suppliers to simplify effective search, controlling their online actions) so, according to that, necessity becomes urgent to develop a firm big data policy, which will take into account the sensitivity of consumer information.

iv. Elaheh Yadegaridehkodi, Mehrbakhsh Nilashi, Liyana Shuib, Mohd Hairul Nizam Bin Md Nasir, Shahla Asadi, Sarminah Samad, Nor Fatimah Awang (2019). Impact of Big Data Firm Performance in Hotel Industry. Electronic Commerce Research and Applications [5].

This study proposes integrating four organizational dimensions (Technology, human, organization, and environment) to identify key factors and how they will affect the firm's performance. The result presents that I.T. expertise was the only critical factor under the human dimension. At the same time, external pressure was stated as an essential factor in the environmental dimension. Also, management support is an essential factor related to the organization dimension, and technology is the most influential dimension. Moreover, the result demonstrates two concerns as the biggest hurdle in the adoption of big data. Those concerns were security and privacy. This outcome will improve the whole business and government's decision in concern of big data adoption.

1.8 METHODOLOGY

In scientific research, it is possible to use a qualitative method to perceive a specific phenomenon, a specific event, study value, and meaning.

And using a quantitative method when there is previous knowledge about a research topic contributes to numerical data and usable statistics.

According to the usability of qualitative and quantitative methods, the research will use the qualitative method.

a) First of all, gather data using a Questionnaire.
b) Analysis of collected data by using SPSS.
c) Result documentation.

According, to a used technique with qualitative methods, will produce trusted and actual results. All of that will give clear evidence of effective and efficient techniques.

1.9 CONTRIBUTION AND IMPACT

A contribution and the effects of this scientific research are to develop knowledge and perception of a scientific attitude. This research will be addressed to get the benefit of using big data and data interoperability, which will positively affect this scientific area. In addition, an innovation aspect that can gain from research ideas and contribution gives a wide area for scientific and innovation of big data speciality.

Recently, Big data concerns and research have become more interested. Especially big data analytics. According to this, the research helps to reduce a shortage in scientific studies in this field and solve problems. From this point, the research discusses the impact of big data features in the context of data interoperability. All of the previous used to create a unique conceptual framework that helps to achieve data interoperability.

The original contribution of research can be as follows:

a) Scientifical (Increase Knowledge productivity in this area).
b) Technological (Different techniques achieved).
c) Theoretical (Supporting the scientific research and using his Narration and inference).
d) Methodological (Use a scientific research methodology as the research pathway).
e) Practical (Gathering data and analyzing it to measure big data features' impact on data interoperability).

The research has a short-term, which represented as a goal and deducted the informatic evidence which helps in big data areas and applications.

Therefore, in the long term, it is possible to develop this orientation. This research or framework will be a part of the integrated system to issue expected provisions based on the inferred information.

They are a research concern regarding the economic and commercial aspects because they affect data interoperability, affecting the community's economic life cycle. Moreover, it is possible to apply this research idea to some business and management projects.

Here, it is necessary to talk about the risks that are related to the research topic:

a) Regulations and rules: they are negatively affected by reliability and sustainability.
b) Interoperability barriers (Conceptual, technological, and organizational).

i. Conceptual barriers (Semantic and syntactic).
ii. Technological barriers (Data source diversity, different platforms, data integration, scheme conflicts and data format, incompatible communication protocol, etc.).
iii. Organizational barriers (responsibility, authority, authorization, personal data protection, etc.).
iv. Data quality and reliability concerns.

1.10 WORK PLAN

a. List of work packages.

WP-1: prepare chapter1 with documentation: (Duration:1 month)

Effect on success: better understanding of Initial thesis issues.

WP-2: prepare chapter2 with documentation: (Duration:1 month)

Effect on success: Theoretical issues of study.

WP-3: Gathering data: (Duration:3 months)

Effect on success: collecting data and choosing an appropriate environment effectively get better results.

WP-4: Analyzing the data: (Duration:2 months)

Effect on success: will be flexible to give the expected results.

WP-5: Results and Chapter3 documentation: (Duration:1 month)

Effect on success: results documentation to be ready for the next stage.

WP-6: Chapter4 documentation and the study conclusion: (Duration: 1 month).

b. Work-Time Schedule.

Table 1.1: Work-Time Schedule

W.P. /No	**W.P. Name**	**Month**								
		1	2	3	4	5	6	7	8	6
1	Prepare chapter1 with documentation.	x								
2	Prepare chapter2 with documentation.		x							
3	Gathering data.			x	x	x				
4	Analyzing the data.						x	x		
5	Results and Chapter3 documentation.								x	
6	Chapter4 documentation and study Conclusion.									x

2. BIG DATA AND INTEROPERABILITY

After the rapid progress in the last few decades and the evolution of the internet, which causes the whole world to be connected and become in the same frequency, a lot of topics are coming to be the surface in a healthy way, such as "Big Data." therefore "Big Data" has been at the forefront of researchers' hot topics, because of the technology evolution in different domains such as industry, social media, healthcare, online trade…ETC, which produced a massive amount of data, makes the researchers think about how we can handle this massive amount of data is produced every day. For instance, communities such as Facebook, Google, and Twitter generate more than 500 terabytes[7] of data a day. And this is a lot of unused data each day. Moreover, we can say that Big Data is also data, but on a vast scale. A vast collection of data that develops rapidly over time is referred to as Big Data. In a nutshell, such data is massive and complicated. As a result, standard data management systems are unable to store or process it effectively. [7]. We can't deal with 'big data' like traditional data; there are many problems that face us, like the rapid growth and expansion of computing systems. We can say that the central issue is the variety of data types. Big data came in three forms:

a) Structured

Data that can be stored, processed in a fixed format, and a table in a school database containing a student's information.

b) Unstructured

Any data with a considerable size and unknown format also challenges communities or organizations to create any value. We take as an example the result of 'Google Search.

c) Semi-structured

Semi-structured data contains two types, ' structured' and 'unstructured,' for example, the data symbolized in an XML file is semi-structured big data.

We can say "Big Data" is different from the usual data that is used before this brings us to the 3V volume, velocity, and variety, these are keys to comprehending how we can measure big data, this makes us relies on that "Big Data" is an area of concern for private and public institutions at every level. And can benefit from technologies derived from big data to make

public or private services provided to citizens and enterprises efficiently, improve policymaking through data-driven decisions, and avoid fraud.

After we learned a few basic things about Big Data, we reached the main challenges of data interoperability in a big data environment and the ability to exchange information between systems. Thus, what is data interoperability? And the relation between it and big data, these questions and more will be posted, and we will try to answer them through a study that will be applied to the ALCPO specialized in the customer's protection area, also will present some of the interoperability challenges that the institution will face in the decision-making process.

2.1 BIG DATA CONCEPT

Big data can be defined as the following:[2]

a) Data with some features such as high volume, velocity, variety, value, and veracity could give organizations an advantage in the marketplace.

b) A complex, extensive Data that couldn't be managed by traditional techniques, but with advanced techniques of NoSQL data management, analysis and visualization are required.

c) A collection of structured, unstructured, and semi-structured data can be used from analytic and predictive machine learning models.

2.1.1 Big Data Types

a) Structured Data: refers to organized tabled Data that can be stored, processed, and retrieved according to a predefined schema. An instance of structured data is a relational database.

b) Unstructured Data: Unlike structured data, unstructured data is not a predefined data model, and it has different types, including images, text, emails, PPT files, transactions, audio, videos…. etc.

c) Semi-structured Data: it is between structured and unstructured. Semi-structured data is not organized in a relational database, but its properties make it easy to analyze and process.

Table 2.1: Shows The Differences Between Big Data Types[8]

Properties	Structured Data	Semi-structure Data	Unstructured Data
Technology	It is based on a relational database table.	Based on XML.	Based on the text and binary data.
Transactions	Matured transactions and various synchronize techniques	DBMS transaction is not matured.	No transaction management.
management	Managed tuples, rows, and tables.	Managed over tuples or graphs is possible	Managed as a whole
Flexibility	Rely on the Schema and be less flexible.	More flexible than structured data.	No schema, and more flexible.
Scalability	Not easy to scale the D.B. schema.	scaling is simple than structured data	more scalable
Query performance	Structured queries allow complex joining	Queries over anonymous nodes are possible	Only textual queries are possible

2.1.2 Big Data Characteristics

Things need to be defined and organized; big data can be organized by illustrating its characteristics to simplify our understanding that characteristics can give many opportunities [9]:

a) Deal with tremendous data.

b) A convenient speed and acceptable time frame can give data value with real-time response.

c) Volume: Huge amount of data is the essential aspect of big data. Data is continuously generated from different sources with various formats and platforms stored in data warehouses to be processed and managed by effective tools like Hadoop.

d) Value: Increasing organization productivity, business process value, and revenues by improving data management tools, and analytics, has become an essential concern for many institutions. This concern could give opportunity and value to the produced data.

e) Veracity: it is about accuracy and reliability by removing uncertain and incomplete data.

f) Velocity: The acceleration in which the data can be created in real-time, and a massive flow from different sources, networks, social media, mobiles, different devices, and machines. This acceleration of this enormous amount of data can make opportunities in concern of valuable decisions and strategic plans competitive.

g) Variety: Refers to the different kinds of big data (structured, unstructured, semi-structured). This data is gathered from numerous sources.

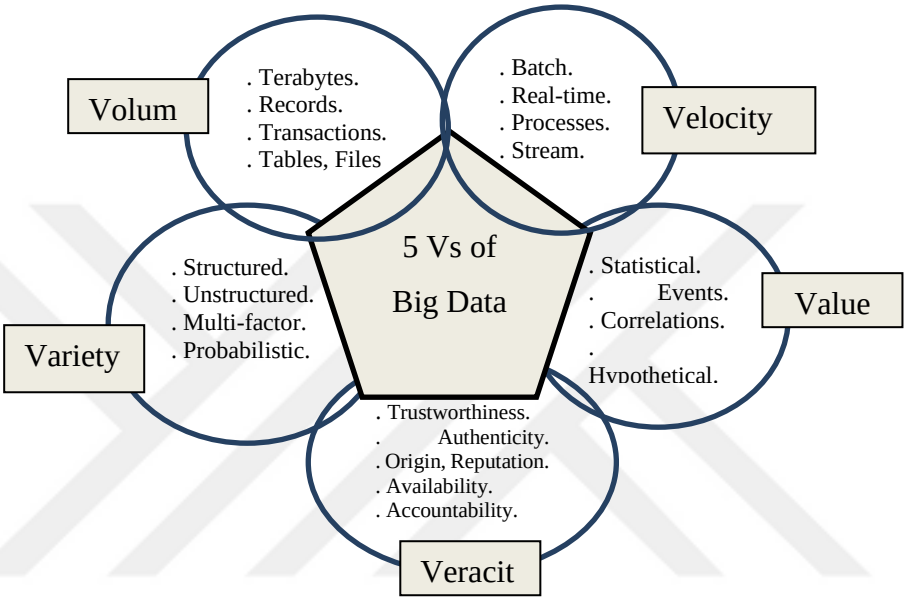

Figure 2.1: The Five Features Of Big Data [9]

Another five big features [10]

1. Big Intelligence refers to the many techniques, theories, and tools. This simulates human intelligence to develop a big data-driven intelligence concept that supports business functions, management, and data analytics.
2. Big Analytics is about gathering, formulating, and analyzing massive data to discover a related pattern and knowledge. That will support decision-makers.
3. Big Infrastructure is concerned with related techniques, platforms, communication, and procedures. All facilities serve big data arguments. Those arguments, including collecting, storing, and analyzing, can be beneficial when using data mining techniques and discovering patterns. For instance, Hadoop is considered a virtual infrastructure environment for big data.

4. Big Service: services mean providing many packages with special functions to serve millions of users worldwide. Big data services are services based on big data. For instance, cloud services and mobile services.
5. Big Market: This indicates the big data-driven markets with all supported concepts (platforms, tools, and services).

2.1.3 Big Data Sources

Data can be existing in many forms. Also, big data, with its massive spread and various forms, could come from the following sources [11]:

a) Media platforms like YouTube, search engines like google and social media (Facebook, Twitter, etc.), and other universal media like images represent enormous user interactions.

b) Cloud storage supports structured and unstructured data and contributes real-time information with many advantages and challenges:

i. Improve flexibility, and scalability and facilitate connectivity and collaboration.
ii. Encourage innovation and competitive positions.
iii. Enable Disaster recovery.
iv. Less control of security, privacy, and latency concerns.

c) The public websites provide a vast amount of diverse information with rapid growth.
d) IoT devices, machines, and different sensors provide a vast amount of real-time information.
e) Databases and data warehouses, modern business process requirements, and business analytics depend on a combination of various resources of big data.

2.2 INTEROPERABILITY CONCEPT

Interoperability as a concept has been defined by many studies but still not participated by all communities. So, according to the IEEE, Interoperability is known as the capability of different heterogeneous systems or components to interact with data and achieve benefits [12]. Another definition introduced interoperability as the ability of organizations to

collaborate and share information [7]. Also, another exciting definition was conducted by (Olaronke Iroju and others); they represent interoperability as the ability of three related topics (information, communications, and software applications) to exchange data. But this exchangeability will be ineffective, consistent, and accurate manner. And then use that arising information in different operations[13]. Here is a concept of different interoperability levels are appeared :[14]

a) No interoperability: stand-alone devices with fully human operation.

b) Technical Interoperability: The possibility to connect devices and make them connected. Technical interoperability can be achieved using a low-level network communication protocol

c) Syntactic interoperability: understanding the data format on both exchangeable devices.

d) Semantic interoperability: concerns with understanding the meaning of the data.

e) Pragmatic Interoperability: Data context and associated data have to be recognized.

f) Dynamic interoperability: it describes the situation of each interoperating device by a state-model concept, which means the states of every device at a given time.

g) Conceptual interoperability: create a complete understanding of the data models, concepts, and standards.

2.3 DATA AND BIG DATA INTEROPERABILITY CONCEPT

Data interoperability concerns the understanding of the meaning of information between the sender and the recipient. So, data interoperability offers a suitable data description across systems.[13]

According to that, a good practice of big data interoperability layers in detail can be conducted as follows [7]:

a) Legal interoperability: in which organizations with various policies and strategies can work together, the legal interoperability could be legislation affected or issues related when manipulating personal data.

b) Organizational interoperability: to make different organizations communicate with each other's, taking into account different barriers. Here, breaking down data silos and using open accessibility will increase overall organizational efficiency and effectiveness. Also, dealing with various data types is a crucial concept in this part. Finally, aligning data requirements to be suitable and support the business process.

c) Semantic interoperability: heterogeneous data sources are an enormous challenge to applying semantic interoperability. This can be done by converging the semantic gap between datasets, mapping semi-automated meanings, and choosing the correct format.

d) Technical interoperability: dealing with data integration patterns and multiple data sources.

2.4 LEVELS OF INFORMATION SYSTEMS INTEROPERABILITY

a) Technical Interoperability: refers to the infrastructure of equipment and software components, platforms, and communications protocols [15].
b) Syntactic Interoperability: Indicates to standardized the data format to improve the exchangeability [15].
c) Organizational Interoperability: is about transferring meaningful data and using it with various information systems with different infrastructures [15].
d) Semantic Interoperability: pertains to the content definition and understanding **of** various data similarly with no ambiguous and inaccurate meaning [15].

2.5 BIG DATA PRIVACY AND SECURITY

Privacy concerns the authority to control how the information has been collected, used, transmitted, and forbidden to become known by unauthorized access.
Security is the practice of protecting information assets during the use, process, and training from unauthorized users, intruders, modifications, cyber-attacks, etc.

Table 2.2: Describes The Difference Between Big data Privacy And Security [16]

S. No	Privacy	Security
1	indicates the appropriate use of customer information and sharing information with the others.	It refers to: Data Confidentiality, availability, and integrity.
2	Safe the user's and organization's rights from unauthorized use.	Protect the overall enterprises in aspect of a security system.
3	Not sharing customer information with a third party without the user's consent.	Uses different techniques like (Encryption, firewall, etc.) to prevent vulnerabilities.
4	privacy cannot be reached without security.	Security can be reached without privacy.

2.5.1 Big Data Security Issues:

a) Distributed Systems: Big data concept built on data and processes in concern of distributed systems, this kind of system is susceptible to be threats [17].

b) Data Access: Illegal access to sensitive data by hackers or unauthorized users is a crucial issue of big data security, in which companies have to control users' identities [17].

c) Incorrect Data: creating fake data and deploying such data inside big data systems to interfere with the system process can seriously impact the results and big data system's performance [17].

d) Privacy Violations: confidentiality and privacy of personal data is a vital issues and threaten sensitive data [17].

e) Inadequate Cloud Security: storing big data without encryption and using intrusion detection systems can be essential for big data security issues.

2.6 INTEROPERABILITY APPROACHES AND ASSESSMENTS

Establishing interoperability has three approaches that have to be relevant to each other: [18]

a) Integrated approach: general format for entire models has to be built. This format has to be admitted by all entities.
b) Unified approach: it refers to a meta-level standard format and provides model mapping.
c) Federated approach: it is about the languages and approach of a task as the standard format.

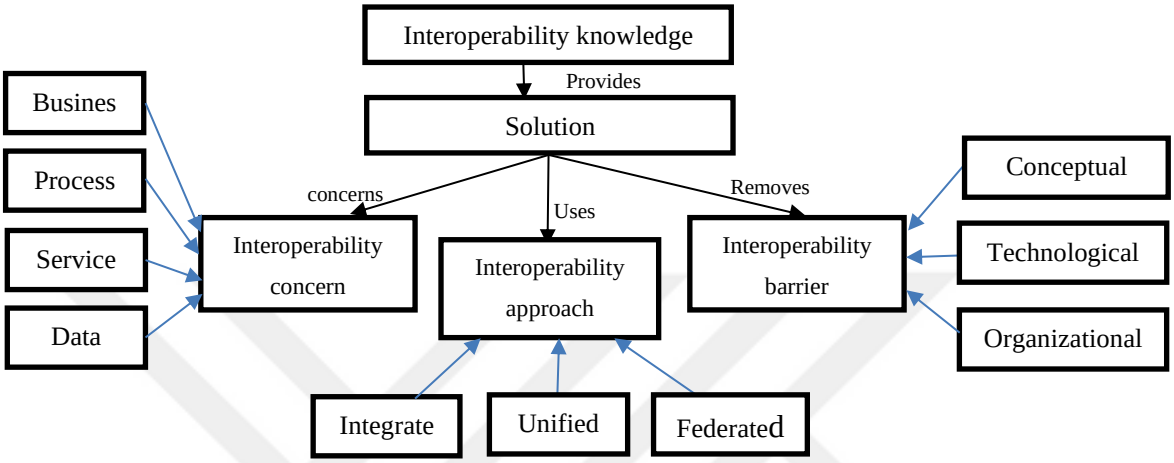

Figure 2.2: Interoperability Basic Concepts And Approaches [18]

2.6.1 INTEROPERABILITY ASSESSMENT:

For better collaboration, the interoperability needs to be improved and measured by defining metrics. That will help the institutions realize their strengths and weaknesses and avoid enterprise interoperability's adverse effects and barriers.

2.6.2 Types Of Assessment:

Three types of interoperability assessment have been considered [19]:

2.6.2.1 The potential assessment

It provides many characteristics of the potential to overcome any hurdles and interoperate with a third partner—Open systems as an example of high potential interoperability. Here in the potentiality interoperability, three barriers (Conceptual (Syntactic/Semantic), Technological, Organizational) are considered according to their impact to evaluate the enterprises. There are five levels of potentiality for each of them: (isolated, initial,

executable, connectable, interoperable) [20]—Moreover, the four levels of interoperability (are business, Processes, services, and data).

a) Isolated: Refers to the interoperability incapacity.
b) Initial: Strong efforts are required for interoperability for a better partner.
c) Executable: executing the interoperability is possible, even if the problem rate is high.
d) Connectable: connect ability with the partner; even the risks can occur with a distant partner.
e) Interoperable: indicates to the evolution of different levels of interoperability.

2.6.2.2 The compatibility assessment

identifying the defeats that may cause problems when the data has to be exchanged. So, the idea is to check the compatibility concerning interoperability barriers. Three concepts have to be conducted to investigate the incompatibility [19].

a) Conceptual Compatibility: could be syntactic or semantic. Syntactic refers to whether the exchanged information has the same syntax or not, and semantic if this information has the same meaning.
b) Organizational Compatibility: authorizes and responsibilities at both sides are considered to check the organization's compatibility.
c) Technological Compatibility: organization I.T. Platform and communication protocols compatibility have to be suitable and acceptable by the partners.

Here we have to deal with many barriers.

Table 2.3. Shows the interoperability barriers that are relevant to the enterprises' concerns [19]

		The barriers to interoperability		
		Conceptual	***Technological***	***Organizational / Legal***
Interoperability Concerns	***Business***	Visions, strategies, Organization cultures, understanding.	- I.T. infrastructure.	- organization structure. - Legislations. - Business rules.
	Process	Syntax and semantics processes	Process interfaces and supporting tools.	Procedures of work, processes organization.
	Service	Semantics to name and describe services.	Interface, architecture.	Responsibility/ authority to manage services.
	Data	- Data representation and semantics. - Data restriction rule.	Data exchange formats.	Responsibility authority to add/delete, change/ update data.

2.6.2.3 The performance assessment

Three criterions have to be considered (interoperability cost, interoperability time, and interoperability quality). Those criterions are evaluated during run-time [20].

The cost: indicates to the cost of updating the application and eliminating the barriers. And it can be (the actual cost and the expected cost).

$$InteropC = (EC - RC)/EC. \quad (1)$$

The time: Refers to the difference between the real requested time and when the information has been exchanged.

$$InteropT = (ET - RT)/ET. \quad (2)$$

The quality: The quality of the information is considered with the performance measurement. So, two types of quality are conducted:

The quality of exchange: Refers to the data that are successfully posted.
The quality of utilization: Refers to the delivered data in comparison with the demanded.

$$InteropQ = AE/TE \quad (3)$$

Conformity: Here, the usability rate of the received information is negotiated.

$$CE = AC / AR \quad (4)$$

2.7 DATA INTEGRATION BASICS AND CONCEPTS

Data integration is a concept of bringing data from diverse data sources into a single and unified vision to give the users chance to interact with one specific system. Integration begins with the absorbing process and includes processes such as data cleaning, ETL process, and transformation. Data integration enables analytics tools to yield efficient, dynamic business intelligence components.

2.7.1 Data Integration Problem:

Integration is a concern of combining multiple information systems, so the users can interact and view a new unified single system, that homogeneous system stated over diverse data sources. Therefore, the data source needs to be expressed using unified semantics and a global model. Since the integration of different data sources that are not fit with each other needs to be combined, so the integration issue depends on [21]:

i. System architecture and components.
ii. The information type (structured, unstructured, semi-structured).
iii. The usability and authority of the information (read-only or read-write).
iv. Resources Availability and accessibility.

Also, heterogeneity has to consider the following:

i. The system hardware and software.
ii. Data schemas, semantics, and models.
iii. Business rules and constraints.

2.7.2 Challenges Of data Integration And Interoperability:

Inadequate resources refer to the lack of the following [22], [23]:

a) Skilled individuals.

b) Financials.

c) Implementation cost.

2.7.3 Integration Strategies:

Five types of integration strategies conducted [21]

a) Manual data integration: selected data has been manually integrated.
b) Middleware data integration: reusable functions are typically used to fix dedicated integration problems.
c) Application-based integration: can access many data sources, integrate their format, and combine the structures into one platform.
d) Standard storage integration: A copy of the data has been stored in a master place and provides a consolidated view.
e) Uniform access integration: This strategy will make the data steady and accessible from various sources and provide uniformity between those different data resources.

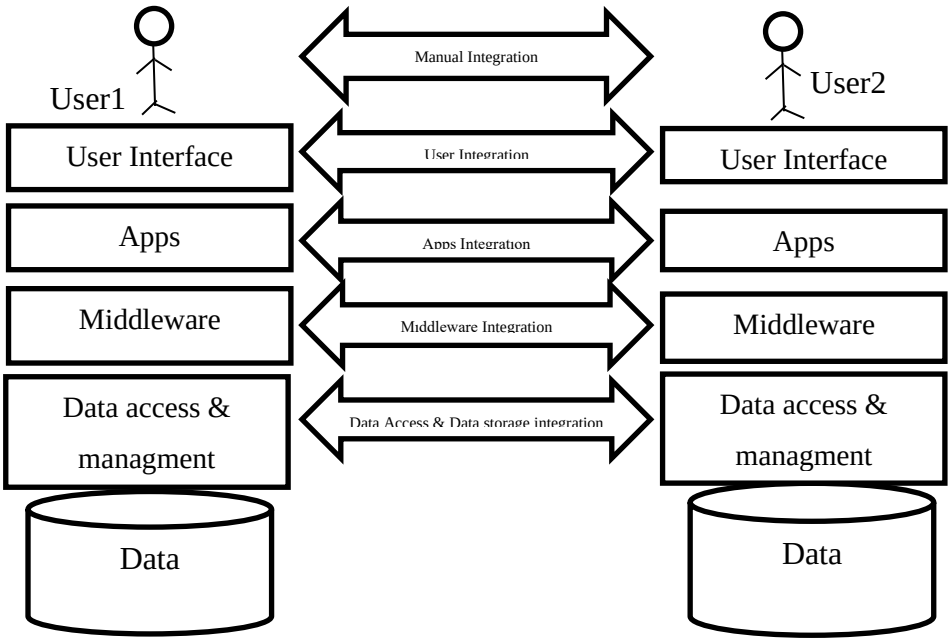

Figure 2.3: Integration Strategies Levels [21]

a) Inconsistency: heterogeneity of the data resources leads to:

1. Inconsistency.

2. more resources to optimize the unstructured, structured, and semi-structured data.

b) Scalability: changes in organizations' systems should pass with many modifications that modify the whole system's efficiency. One of these future expected changes is scalability and transferring from legacy systems to new ones.

c) Data Sources: taking data from different resources, especially external ones, may cause different aspects such as data quality, sharing data, and privacy.

d) Sustainability: This is concerned with keeping integration efforts with best practices and modern techniques.

e) Different Versions Of Data: new systems generally generate data diversity from many sources (IoT devices, sensors, etc.).

f) Query Optimization: Reducing the queries by finding the most efficient way to run queries according to the query plan. Using joins, grouping, aggregations by many sequence query languages (SQL) strings and high-level query optimization can positively affect data processing and storage performance.

g) Providing Support System: organizations need to Establish reporting system to handle errors and updates.

h) Data Quality: many data resources may negatively affect the data quality, which will already affect interoperability.

2.7.4 Data Integration Patterns:

a) Data Consolidation: The ETL tool collects the data from different resources, manipulates them by functional format, and releases them into a new central database. So, reducing the number of storage stations is the goal that can be achieved from data consolidation.

b) Data Propagation: is about the ease of accessibility by copying the data and distributing them to other databases.

c) Data Virtualization: views the data from various data sources and locations in a real-time and integrated interface.

d) Data Federation: EII (Enterprise's information integration) is a data virtualization technique where diverse data of many systems with a unified view and accessibility is arranged in a single virtual database.

e) Data Warehousing: providing queries and analysis optimization by combining a relational database, involving the outlier cleaning, formatting, and saving data.

2.8 BIG DATA AND E-GOVERNMENTS

The tendency of big data and e-government to discover user behaviour patterns becomes a more productive contribution that shows better consumer services, improving business intelligence in creating and predicting business value and decisions. The new e-government strategy adopts big data to better understand different users' emotions, which will help achieve reliable, effective, and intelligent services to cooperate with other organizations and serve their citizens. Moreover, adopting big data has a positive effect on creating value and motivation for the digital innovation of e-government. Creating value has been indicated as the essential principle of big data and e-government, increasing the importance of big data analysis, management, and benefits [24].

In the following, some benefits will be described:

a) different data types are available.

b) Improve data transparency.

c) Enhance data queries and transactions.

d) Providing an efficient resource of big data.

e) Support decision-making and data integration.

f) Increasing an organization's revenues by relying on valuable data and controlling the resources.

g) Raising e-government services and reforming people's life situations.

2.8.1 E-government And Big Data Impediments

Big data for e-government has many challenges; here, the challenges from different perspectives will be conducted:

Technology prospects: E-government and big data utilization are crucial issues to be performed without sufficient infrastructure of ICT; that ICT infrastructure has to be conducted with big data future techniques (data management tools, storage, and analyzing mechanisms), which will increase the value of the meaningful data. So, as refers in previous the technological challenges could be as the following:

a) The ICT equipment's capacity, flexibility, and features of storage and processing compatibility with that hug data.
b) Organization regulations and legislations.
c) Lack of Skilled individuals for big data tools and analytics.
d) Analyzing a significant amount of data, releasing modelling and structure challenges.
e) Data management tools, uses, and mastery.
f) Human prospects: organizations need to know their customers and how retention them. So, the customer's behaviour needs to be tracked to improve the institution's services and share information with a third party. The challenges can come from the people aspects.
g) Business Process prospects: generating a vast amount of data and earning value for enriching services. This will increase the demand for government support to improve decision-making, motivate the competition, and enhance performance. So, that support can be established by creating a comprehensive, effective strategy plan, looking forward to a satisfying partnership, and motivating the e-leadership role inside e-government.

3. DATA ANALYSIS

3.1 TESTS OF NORMALITY.

The main aim of this chapter is to compare the shape of the current sample distribution data to the shape of a standard curve. It is assumed that when the current study sample is normally shaped, it can decide that the population from which the current study sample came is normally shaped. As a result, it can be concluded that normality is assumed to be achieved. As demonstrated by Table 3.1 along with Figure 3.1, "Histogram", it has been revealed that the existing data have normally shaped. In Particular, as seen from Table 3.1, the possibility-value of Kolmogorov-Smirnova 0.096>0.05 and the results of Shapiro-Wilk 0.109> 0.05, as such values indicate that the null hypothesis is confirmed. In Particular, the null hypothesis declares that "the sample is drawn from an ordinary distribution, or it could be drawn from a non-normal distribution [28]. In addition, the histogram or Figure 3.1 informs that the existing sample data are symmetric, and then the existing sample data are typically distributed. Therefore, the existing sample data are not significantly different from the regular population data. To summarize, based on the findings of the normality test table 3.1 and Figure 3.1, the researcher concludes that it is most likely that parametric test statistics (for instance, regression analysis and correlation analysis) can be utilized to test the current study hypotheses.

Table 3.1: Tests of Normality Table

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Data	.096	38	.200*	.953	38	.109

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

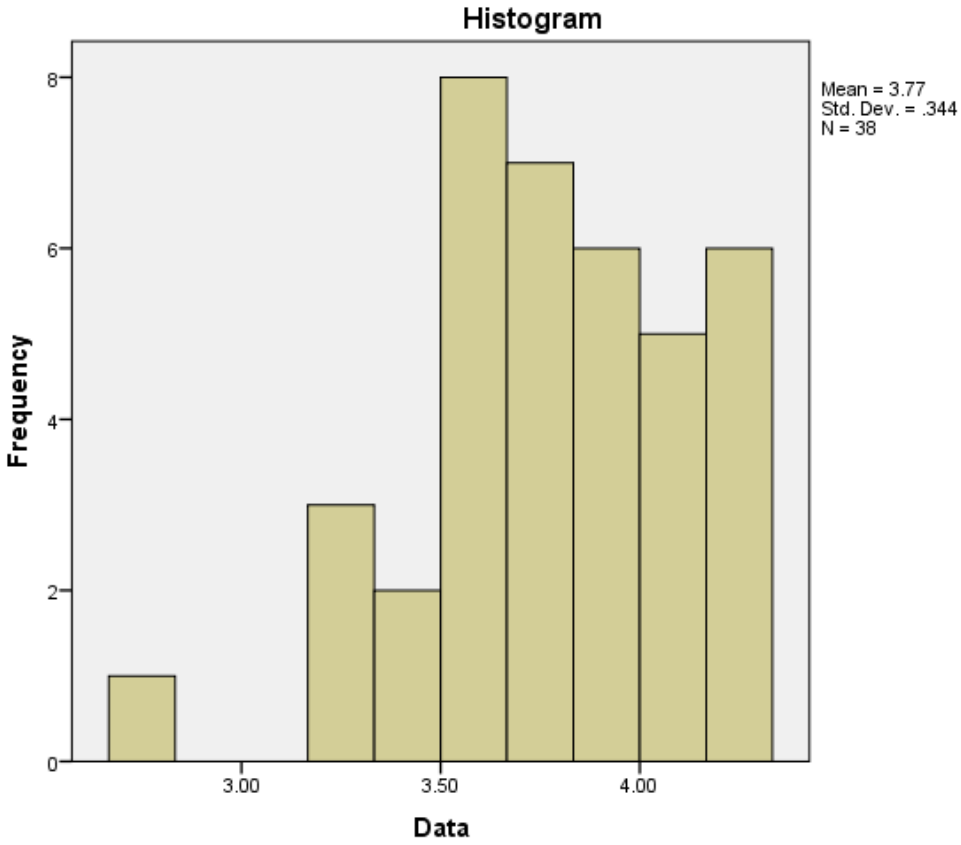

Figure 3.1: Tests of Normality Histogram

3.2 DEMOGRAPHIC PROFILE OF THE RESPONDENTS.

3.2.1 Gender Variable.

Both Table 3.2 and Figure 3.2 reveal the demographic profile of the study's respondents regarding their gender. Based on Table 3.2, it can be observed that the male group of the study participants forms the most significant percentage (89.5%). In comparison, the female group of the study participants forms the smallest percentage (10.5%). Therefore, it can be concluded that the male community is more interested in the study subject.

Table 3.2: Gender Variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	34	89.5	89.5	89.5
	Female	4	10.5	10.5	100.0
	Total	38	100.0	100.0	

Figure 3.2 confirms the findings of the frequency distribution of gender of the respondents, Table 3.2. The respondents' plurality, representing 89.47%, are males, and 10.5% are females.

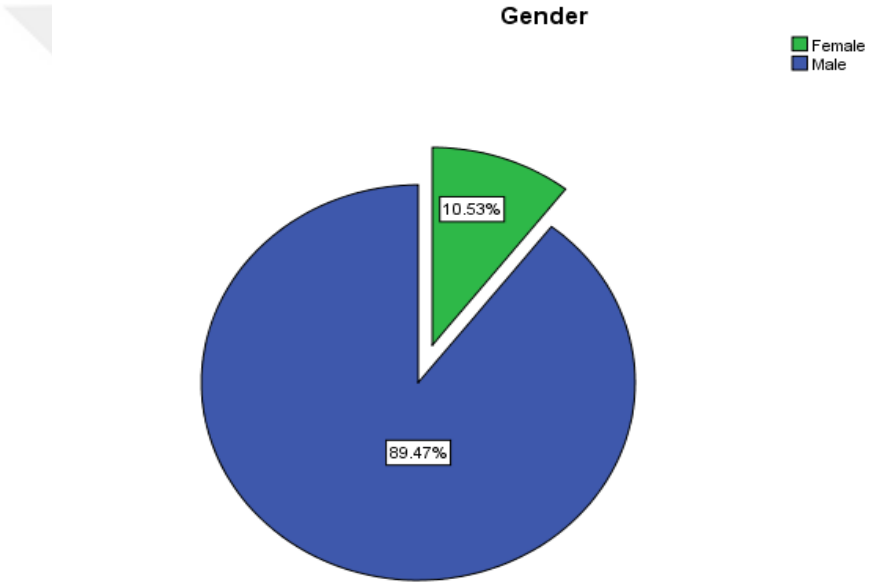

Figure 3.2: Gender Variable

3.2.2 Age Variable.

Both Table 3.3 and Figure 3.3 reveal the demographic profile of the study's respondents about their ages. Based on Table 3.3, it can be observed that the more than 40 years old group of the study participants forms the most significant percentage (44.7%). In contrast, the study participants less than 30 years old group forms the smallest percentage (10.5%).

Therefore, it can be concluded that the more than 40 years old community is more interested in the study subject.

Table 3.3: Age Variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Less than 30 years old	4	10.5	10.5	10.5
	Between 30 to 40 years old	17	44.7	44.7	55.3
	More than 40 years old	17	44.7	44.7	100.0
	Total	38	100.0	100.0	

Figure 6 confirms the findings of the frequency distribution of age of the respondents, Table 3.3. The plurality of the respondents, which represents 44.74%, is more than 40 years old.

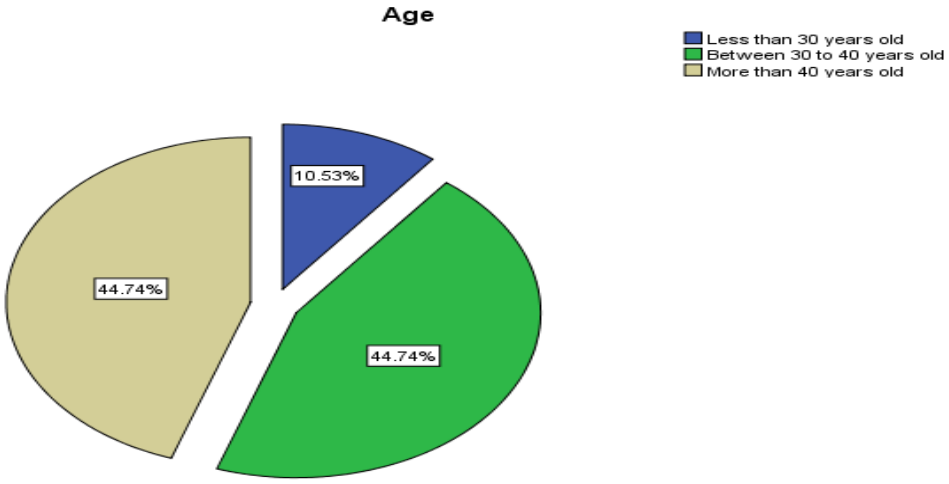

Figure 3.3: Age Variable

3.2.3 Manargial Level Variable.

Both Table 3.4 and Figure 3.4 reveal the demographic profile of the study's respondents regarding their managerial level. Based on Table 3.4, it can be observed that the lower-level management group of the study participants forms the most significant percentage (42.1%).

At the same time, the top-level management group of the study participants forms the smallest percentage (23.7%). Therefore, it can be concluded that the lower-level management community is more interested in the study subject.

Table 3.4: Managerial Level Variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Top Level Management	9	23.7	23.7	23.7
	Middle Level Management	13	34.2	34.2	57.9
	Lower-Level Management	16	42.1	42.1	100.0
	Total	38	100.0	100.0	

Figure 3.4 validates the findings of the respondents' frequency distribution of management levels, Table 3.4. The majority of the respondents, which represents 34.21%, are from lower-level management.

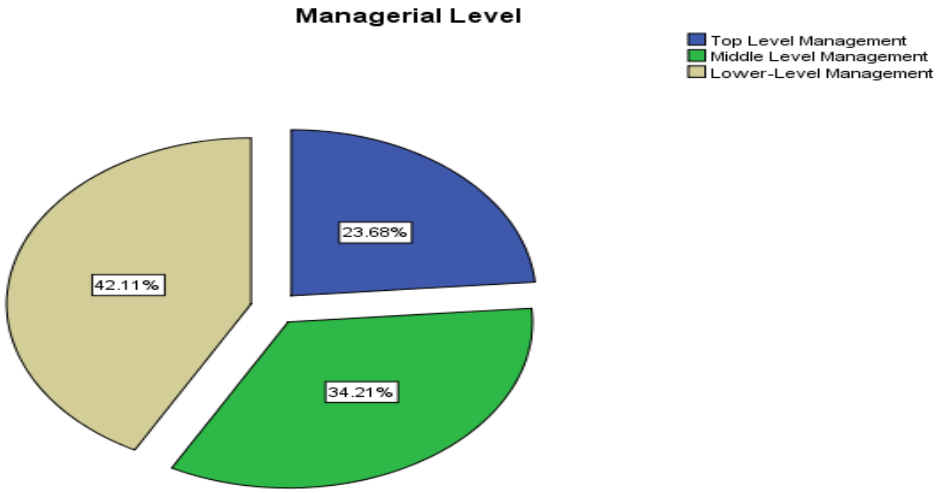

Figure 3.4: Managerial Level Variable

3.3 RELIABILITY OF THE QUESTIONNAIRE

The possibility that a product, system, or service will perform its intended function satisfactorily for a certain period or in a specific environment without fail is defined as reliability [29]. Cronbach's alpha is the most applied to measure the internal consistency of the scales. As a result, it was applied to evaluate the overall reliability of the big data features and the interoperability constructs scale. An alpha value of 0.60 and higher is often the standard for internally consistent established scales. Table 3.5 lists Cronbach's alpha value of all the scale measures of big data features and the interoperability constructs scale, which is over 0.77, thus exhibiting a satisfactory level of construct reliability. Finally, the mean and standard deviation s of the items of variables are considered to be quite reasonable.

Table 3.5: Reliability Scale Analysis

Construct	Mean	Std. Deviation
Value	3.6955	0.47458
Volume	3.8526	0.89887
Variety	3.9060	0.47354
Velocity	3.8737	0.50387
Veracity	3.9263	0.47743
Data Interoperability Potentiality	3.5474	0.57645
Data Interoperability Compatibility	3.6579	0.62445
Data Interoperability Performance	3.5482	0.42076
Cronbach's Alpha = 0.776		

3.4 CORRELATION ANALYSIS

The correlation analysis aims at testing the following main hypothesis:

There is a considerable positive correlation between big data features and data interoperability.

The above main hypothesis is split into the following sub-hypotheses:

I. There is a considerable positive correlation between the data value, data interoperability potentiality, data interoperability compatibility and data interoperability performance.

II. There is a considerable positive correlation between the data volume, data interoperability potentiality, data interoperability compatibility and data interoperability performance.

III. There is a considerable positive correlation between the data variety, data interoperability potentiality, data interoperability compatibility and data interoperability performance.

IV. There is a considerable positive correlation between the data velocity, data interoperability potentiality, data interoperability compatibility and data interoperability performance.

V. There is a considerable positive correlation between data veracity, data interoperability potentiality, data interoperability compatibility, and performance.

Table 3.6 depicts the bivariate correlation matrix. And displays big data features and their relationships with the interoperability variables. The data value variable is positively correlated with data interoperability potentiality ($r = 0.372$, $p<0.05$), data interoperability compatibility ($r = 0.440$, $p<0.01$), data interoperability performance ($r = 0.530$, $p<0.01$). Thus, the first hypothesis is confirmed and supported. The second hypothesis has been revealed that is inadmissible as the correlation is not considerable. That is the data volume is not correlated with data interoperability potentiality ($r = 0.206$, $p>0.05$), data interoperability compatibility ($r = 0.108$, $p>0.05$), data interoperability performance ($r = 0.250$, $p>0.05$). In the third hypothesis, it has been revealed that the data variety is only correlated with Data Interoperability Potentiality ($r = 0.392$, $p<0.05$). The fourth hypothesis shows that the data velocity is only correlated with data interoperability compatibility ($r = 0.417$, $p<0.01$). While in the fifth hypothesis, it has been shown that the data veracity is correlated with data interoperability potentiality ($r = 0.382$, $p<0.05$) and interoperability performance ($r = 0.525$, $p<0.05$).

Table 3.6: Correlation Analysis

Variables	Value	Volume	Variety	Velocity	Veracity	Data Interoperability	Data Interoperability	Data Interoperability
Value	1							
Volume	0.042	1						
Variety	-0.028	.476**	1					
Velocity	0.280	0.113	0.001	1				
Veracity	.341*	.420**	.351*	.419**	1			
Data Interoperability Potentiality	.372*	0.206	.392*	0.192	.382*	1		
Data Interoperability Compatibility	.440**	0.108	-0.089	.417**	0.230	.437**	1	
Data Interoperability Performance	.530**	0.250	0.256	0.136	.525**	.430**	.510**	1
*. Correlation is considerable at the 0.05 level (2-tailed).								
**. Correlation is considerable at the 0.01 level (2-tailed).								

3.5 REGRESSION ANALYSIS

The main aim of regression analysis is to show the significant impact of the big data features on interoperability. The first hypothesis that is tested by the regression analysis is shown below:

There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Potentiality.

By estimating the assumption of no multicollinearity, two or more independent variables in a regression model have a strong correlation, multicollinearity exists. For revealing the multicollinearity problem, there are a few guidelines that can be applied. The first is that if the variance inflation factor (VIF) exceeds 10, there is cause for concern. In addition, if Tolerance below 0.1 indicates a significant problem and Tolerance below 0.2 indicates a potential problem.

According to Table 3.7, the VIF for our current model values is well below ten, while the tolerance statistics are above 0.2; as a result, we can reasonably assume that there is no collinearity in the regression data.

Table 3.7: Multicollinearity Test of Data Interoperability Potentiality Model

Data Interoperability Potentiality		
Variables	**Tolerance**	**VIF**
Value	0.840	1.191
Volume	0.698	1.433
Variety	0.718	1.392
Velocity	0.785	1.273
Veracity	0.590	1.695

When analyzing the first regression model, as in Table 3.8, using R- Squared (R^2) and ANOVA F, it can be comprehended that the value of R Squared (R^2=0.32), as R^2 points that about 32% of changes in the value of the data interoperability potentiality variable can be attributable to Value, Volume, Variety, Velocity and Veracity. This percentage indicates that the regression model of the current hypothesis has a decent Goodness-of-Fit. The R- Squared finding can be confirmed by looking at the value of ANOVA F, which is (F =3.007), which is considerable at 0.05. Therefore, the current researcher might infer that the regression model of the current hypothesis has a decent Goodness-of-Fit. Based on Table 3.8, the first hypothesis is not supported, as the Standardized Coefficients (β) of Value, Volume, Velocity and Veracity (β=0.0.322, p>0.05; β=-0.-0.052, p>0.05, β=0.048, p>0.01 and β=0.142, p>0.01) respectively are non-significance. While the Standardized Coefficients (β) of Variety are considerable (β=0.375, p<0.05). That is, only data Variety has a positive effect on the data interoperability potentiality variable.

Table 3.8: Regression Analysis of Data Interoperability Potentiality

Independent Variable	Dependent Variable: Data Interoperability Potentiality
Value	0.322
Volume	-0.052
Variety	0.375*
Velocity	0.048
Veracity	0.142

Table 3.8: Regression Analysis of Data Interoperability Potentiality (table continued)

R2	0.32
ANOVA F	3.007*
* Standardized Coefficient is significant at the 0.05 level	
** Standardized Coefficient is significant at the 0.01 level, n=38	

There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Compatibility.

According to Table 3.9, because the VIF values in our present model are considerably below ten and the tolerance statistics are over 0.2, we can reasonably assume that there is no collinearity in the regression data.

Table 3.9: Multicollinearity Test of Data Interoperability Compatibility Model

Data Interoperability Compatibility		
Variables	Tolerance	VIF
Value	0.840	1.191
Volume	0.698	1.433
Variety	0.718	1.392
Velocity	0.785	1.273
Veracity	0.590	1.695

To evaluate the second regression model, as demonstrated in Table 3.10, utilizing R-Squared (R2) and ANOVA F, it can be recognized that the value of R Squared (R2=0.306), as R2 marks that about 31% of the shift in the value of the data interoperability compatibility variable can be attributable to Value, Volume, Variety, Velocity and Veracity. This percentage indicates that the regression model for the current hypothesis has a decent Goodness-of-Fit. The R-squared finding can be backed up by looking at the ANOVA F value (F =2.825), which is considerable at 0.05. Therefore, the current researcher might infer that the regression model of the current hypothesis has a decent Goodness-of-Fit. According to Table 3.10, the second hypothesis is not entirely supported, as the Standardized Coefficients (β) of Volume, Variety, Velocity and Veracity (β=0.0.133, p>0.05; β=-0.-0.131, p>0.05, β=0.317, p>0.01 and β=-0.033, p>0.01) respectively are non-significance. At the same time, the Standardized Coefficients (β) of the vale are considerable (β=0.353, p<0.05). That is, only data value has a positive effect on the data interoperability compatibility variable.

Table 3.10: Regression Analysis of Data Interoperability Compatibility

Independent Variable	Dependent Variable: Data Interoperability Compatibility
Value	0.353*
Volume	0.133
Variety	-0.131
Velocity	0.317
Veracity	-0.033
R2	0.306
ANOVA F	2.825*
* Standardized Coefficient is considerable at the 0.05 level	
** Standardized Coefficient is considerable at the 0.01 level, n=38	

There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Performance. According to Table 3.11, we can reasonably infer no collinearity within the regression data since the VIF values are substantially below ten and the tolerance statistics are over 0.2 for our current model.

Table 3.11: Multicollinearity Test of Data Interoperability Performance Model

Data Interoperability Performance		
Variables	Tolerance	VIF
Value	0.840	1.191
Volume	0.698	1.433
Variety	0.718	1.392
Velocity	0.785	1.273
Veracity	0.590	1.695

The value of R^2(R2=0.454) indicates that about 45% of changes in the value of the data interoperability performance variable can be credited to Value, Volume, Variety, Velocity, and Veracity when evaluating the third regression model, as seen in Table 3.12, by using R-squared(R^2) and ANOVA F. This percentage implies that the model used in the given hypothesis has a good Goodness-of-Fit. By investigating the value of ANOVA F, the R-squared result can be confirmed. The R-squared result can be validated by looking at the ANOVA F value (F =5.305), which is notable at 0.05. As a result, the current researcher might infer that the regression model of the current hypothesis has a decent Goodness-of-Fit. The third hypothesis is not fully supported Based on Table 3.12, as the Standardized

Coefficients (β) of Volume, Variety and Velocity (β=0.032, p>0.05; β=-0.119, p>0.05, β=-0.152, p>0.01) are all non-significance. Simultaneously, the Standardized Coefficients (β) of value and veracity are significant (β=0.444, p<0.01 and β=0.382, p<0.05). Furthermore, data value and veracity have a positive impact on the data interoperability performance characteristic.

Table 3.12: Regression Analysis Of Data Interoperability Performance

Independent Variable	Dependent Variable: Data Interoperability Performance
Value	0.444**
Volume	0.032
Variety	0.119
Velocity	-0.152
Veracity	0.382*
R2	0.454
ANOVA F	5.305*
* Standardized Coefficient is considerable at the 0.05 level	
** Standardized Coefficient is considerable at the 0.01 level, n=38	

3.6 SUMMARY OF HYPOTHESIS TEST

The bellow table showed the summarizing findings of the regression analysis concerning the study hypotheses.

Table 3.13: Summary Of Hypothesis Test

There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Potentiality	
Independent Variable	Dependent Variable: Data Interoperability Potentiality
Value	Not supported
Volume	Not supported
Variety	**Supported**
Velocity	Not supported

Table 3.13: Summary Of Hypothesis Test (table continued)

Veracity	Not supported
There is a significant effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Compatibility	
Independent Variable	Dependent Variable: Data Interoperability Compatibility
Value	**Supported**
Volume	Not supported
Variety	Not supported
Velocity	Not supported
Veracity	Not supported
There is a significant effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Performance	
Independent Variable	Dependent Variable: Data Interoperability Performance
Value	**Supported**
Volume	Not supported
Variety	Not supported
Velocity	Not supported
Veracity	**Supported**

3.7 LITERATUE COMPARSION

Here a comparsion of many studies have been conducted as follows:

Table 3.14: Literature Comparison

Study title	Study author	Study conclusion
Assessing the impact of big data on firm innovation performance: Big Data is not always better data (2020).	Maryam Ghasemaghaei, Goran Calic	The study investigates the impact of three big data features (volume, velocity, and variety) on innovation performance (efficiency and efficacy). The study implies that: 1- velocity and variety positively affect a firm's innovation performance. 2- data velocity has a substantial role than the others. 3- data volume not playing that significant role. And it is crucial to negotiate big data features as separate concerns, which will enhance overall innovation performance.
Understanding the impact of big data on firm performance: the necessity of conceptually differentiating among big data characteristics (2020).	Maryam Ghasemaghaei	This study reveals the impact of the essential big data characteristics (Volume, Velocity, and Variety) on firm performance and the mediating role of data veracity and value to stand that relation. The findings present that: 1- variety improves data value, whereas the volume and velocity do not impact. 2- veracity has a negative effect on data volume. 3- variety and velocity are affected positively.
Big data's impact on	Nir Kshetri.	The study investigates the relation between big data characteristics, privacy, security, and consumer welfare.

Table 3.14: Literature Comparison (table continued)

privacy, security, and consumer welfare (2014)		This relation is examined from different sides (data collection, storing, sharing, and accessibility). Using big data and specific features and linking them with gathering, storing, manipulating, and being accessible by another partner may lead to being stuck with many privacy and security issues. The study concludes that: 1- the consumers will not be comfortable as companies know and use more than they intentionally provide. 2- unskilled and technologically inexperienced consumers negatively affect consumer welfare.
The impact of Big Data on Firm Performance in Hotel Industry. Electronic Commerce Research and Applications.	Elaheh Yadegaridehkodi, Mehrbakhsh Nilashi, Liyana Shuib, Mohd Hairul Nizam Bin Md Nasir, Shahla Asadi, Sarminah Samad, Nor Fatimah Awang	This study aims to propose integrating four organizational dimensions (Technology, human, organization, and environment) to identify key factors and how they will affect the firm's performance. The result presents that: 1- IT expertise was the only critical factor under the human dimension. 2- External pressure was stated as an essential factor in the environmental dimension. 3- Management support is an essential factor related to the organization dimension. 1- Technology is the most influential dimension. 5- The result demonstrates two concerns as the biggest hurdle in the adoption of big data. Those concerns were security and privacy. This outcome will improve the whole business and government's decision in concern of big data adoption.

Table 3.14: Literature Comparison (table continued)

Current study Tha impact of big data features on interoperability concepts.	Sana Ibrahim Shatwan (2021).	The study conducts the impact of big data features (value, volume, variety, velocity, and veracity) on big data interoperability (potential interoperability, compatible interoperability, performance interoperability). And the following have been concluded: I. Data value variable is positively correlated with data interoperability (potentiality,compatibility and performance) and has positive effect on data interoperability (compatibility, performance). II. There is no significant correlation and no effect between the data interoperability measurements(interoperability potentiality, data interoperability compatibility, data interoperability performance) and data volume. III. Data variety is the only variable strongly correlated and has positive effect on the Data interoperability Potentiality, and there is no correlation with the others. IV. Data veracity variable correlated with interoperability potentiality and interoperability performance and not correlated to the data interoperability compatibility, also has positive effect on data interoperability performance. V. Data velocity is strongly correlated with interoperability compatibility but not correlated with interoperability potentiality and performance, and there is no positive effect.

4. CONCLUSION AND RECOMMENDATIONS

4.1 CONCLUSION

We can conclude that not all of the big data features had a positive correlation between them and data interoperability potentiality, compatibility and performance, as the study proofed that only data value variable has a positive correlation on all of big data interoperability measurements, while it has positive effect on data interoperability compatibility and performance, data variety variable has a positive correlation and a positive effect with only data interoperability potentiality, as for data velocity variable has appositive correlation with only data interoperability compatibility and has no effect on big data interoperability measurements, and data veracity variable has appositive correlation with data interoperability potentiality and performance and has a positive effect only on data interoperability performance, while data volume variable hos no correlation or effect on big data interoperability measurements ,this has a big effect on how the organization could get befits from their big data to develop their work inside and outside their foundation.

from the results of the study, the following conclusion can be conducted:

I. Some Big data features do not correlate with big data interoperability measurements but positively affect them.

II. Some of the Big data features are correlated with big data interoperability measurements but affect them.

4.2 RECOMMENDATIONS

I. organizations need to be concerned with the value of the data. Sharing this data with another potential partner will positively affect the three data interoperability measurements and the whole system's efficiency.

II. on the contrary, data volume is not an essential feature if it is not associated with other characteristics such as value, veracity, velocity, and variety.

III. also, taking care with the variety, and velocity to achieve the best level of data interoperability.

IV. subsequently, achieving a high level of data interoperability means better administrative performance and more accurate decisions

REFERENCES

[1] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.

[2] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Libr. Rev.*, vol. 65, no. 3, pp. 122–135, 2016, doi: 10.1108/LR-06-2015-0061.

[3] M. Ghasemaghaei and G. Calic, "Assessing the impact of big data on firm innovation performance: Big data is not always better data," *J. Bus. Res.*, vol. 108, no. April 2019, pp. 147–162, 2020, doi: 10.1016/j.jbusres.2019.09.062.

[4] N. Kshetri, "Big datas impact on privacy, security and consumer welfare," *Telecomm. Policy*, vol. 38, no. 11, pp. 1134–1145, 2014, doi: 10.1016/j.telpol.2014.10.002.

[5] E. Yadegaridehkordi *et al.*, "The impact of big data on firm performance in hotel industry," *Electron. Commer. Res. Appl.*, vol. 40, 2020, doi: 10.1016/j.elerap.2019.100921.

[6] M. Ghasemaghaei, "Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics," *Int. J. Inf. Manage.*, vol. 57, no. August, 2021, doi: 10.1016/j.ijinfomgt.2019.102055.

[7] "SC508DI07171 D05 . 02 Big Data Interoperability Analysis Document Metadata."

[8] A. De Mauro, M. Greco, and M. Grimaldi, "What is big data? A consensual definition and a review of key research topics," *AIP Conf. Proc.*, vol. 1644, pp. 97–104, 2015, doi: 10.1063/1.4907823.

[9] M. Sowmya and N. Sravanthi, "Big Data: An Overview of Features, Tools, Techniques and Applications," *Int. J. Eng. Sci. Comput.*, vol. 7, no. 6, pp. 13644–13647, 2017.

[10] Z. Sun, "10 Bigs : Big Data and Its Ten Big Characteristics," *Manag. Perspect. Intell. Big Data Anal.*, no. January, 2018.

[11] “Thorn Technologies’ cloud computing and enterprise software blog.” https://www.thorntech.com/blog/ (accessed May 07, 2021).

[12] P. Pagano, L. Candela, and D. Castelli, “Data interoperability,” *Data Sci. J.*, vol. 12, no. July, pp. 19–25, 2013, doi: 10.2481/dsj.GRDI-004.

[13] O. Iroju, A. Soriyan, I. Gambo, and J. Olaleke, “Interoperability in Healthcare: Benefits, Challenges and Resolutions,” *Int. J. Innov. Appl. Stud.*, vol. 3, no. 1, pp. 262–270, 2013, [Online]. Available: http://www.ijias.issr-journals.org/abstract.php?article=IJIAS-13-090-01.

[14] M. Robkin, S. Weininger, B. Preciado, and J. Goldman, “Levels of conceptual interoperability model for healthcare framework for safe medical device interoperability,” *ISPCE 2015 - IEEE Symp. Prod. Compliance Eng. Proc.*, 2015, doi: 10.1109/ISPCE.2015.7138703.

[15] R. Rezaei, T. K. Chiew, and S. P. Lee, “A review of interoperability assessment models,” *J. Zhejiang Univ. Sci. C*, vol. 14, no. 9, pp. 663–681, 2013, doi: 10.1631/jzus.C1300013.

[16] P. Jain, M. Gyanchandani, and N. Khare, “Big data privacy: a technological perspective and review,” *J. Big Data*, vol. 3, no. 1, 2016, doi: 10.1186/s40537-016-0059-y.

[17] “5 types of big data security issues | Artificial Intelligence |.” https://www.allerin.com/blog/5-types-of-big-data-security-issues (accessed May 07, 2021).

[18] D. Chen, B. Vallespir, and N. Daclin, “An approach for enterprise interoperability measurement,” *CEUR Workshop Proc.*, vol. 341, pp. 1–12, 2008.

[19] G. da Silva Serapião Leal, W. Guédria, and H. Panetto, “Interoperability assessment: A systematic literature review,” *Comput. Ind.*, vol. 106, pp. 111–132, 2019, doi: 10.1016/j.compind.2019.01.002.

[20] B. Of, *ICMI-2021*. 2021.

[21] P. Ziegler and K. R. Dittrich, “Data Integration — Problems, Approaches, and Perspectives,” *Concept. Model. Inf. Syst. Eng.*, pp. 39–58, 2007, doi: 10.1007/978-3-540-72677-7_3.

[22] A. Kadadi, R. Agrawal, C. Nyamful, and R. Atiq, “Challenges of data integration and interoperability in big data,” *Proc. - 2014 IEEE Int. Conf. Big Data, IEEE Big Data 2014*, pp. 38–40, 2014, doi: 10.1109/BigData.2014.7004486.

[23] “The Fundamentals of Data Integration - DATAVERSITY.” https://www.dataversity.net/the-fundamentals-of-data-integration/ (accessed May 07, 2021).

[24] Z. A. Al-Sai and L. M. Abualigah, “Big data and E-government: A review,” *ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc.*, no. May, pp. 580–587, 2017, doi: 10.1109/ICITECH.2017.8080062.

[26] Field, A. P. (2000). "Discovering Statistics Using SPSS for Windows: Advanced Techniques for the Beginner: Sage Publications,".

[27] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2011). "Multivariate Data Analysis. New Jersey: PrenticeHall: Inc,".

[28] Ruppert, D. (2014). "Statistics and Finance: An Introduction: Springer New York,".

[29] Sekaran, U., & Bougie, R. (2016). "Research methods for business: A skill building approach: John Wiley & Sons,".

APPENDIX A

Thesis Questionnaire

MEASURING THE IMPACT OF BIG DATA CHARACTERISTICS ON DATA INTEROPERABILITY

We hope that you will kindly answer the following questionnaire, which raises questions about the impact of the different characteristics of big data on the process of exchanging data between different institutions and departments within the same organization.

Sex

- Male
- Female

Age

- less than thirty years
- between thirty and forty years
- more than forty years

vocation level

- higher management
- middle management
- Executive management

The data that is exchanged within the organization is with a high value

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The exchanged data has a positive effect on decision-making

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The exchanged data meets the required needs

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The value of the information has a positive effect on the exchange of data

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Received data is same as required data

- Agree
- Strongly agree
- Disagree
- Strongly disagree

There is no point in exchanging data that has no value

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The data exchanged meets the required needs

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The size of the transmitted data is larger than required

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The volume of data negatively affects the ease and speed of its exchange

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The size of the data negatively affects its quality

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The volume of data exchanged negatively affects the decision-making process

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Due to the volume of the exchanged data some technical problems occur

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Is the sent data various (structured, unstructured, hybrid) structured: it is formatted within a specific rows and columns data model. For example, unstructured SQL relational databases: unformatted data takes many forms e.g. chat posts, different sensor data, pictures and videos, mail Letter, texts. Hybrid: is between structured and unstructured data that uses metadata, tags, and semantics. Example: XML

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The diversity of data sources positively affects their quality

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The diversity of data sources positively affects their exchange

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The diversity of data sources positively affects the decision-making process

- Agree
- Strongly agree
- Disagree
- Strongly disagree

the diverse data requires different techniques

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Difficulty making use of different source data

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Sharing data from different sources requires high costs

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The required information and data arrive quickly enough when needed

- Agree
- Strongly agree

- Disagree
- Strongly disagree

There is a speed in the response of the equipment and devices with the data sent and received

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The equipment and systems used provide sufficient speed for the exchange of information

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The received data provides the required speed in the administrative decision-making process

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Fast data transfer positively affects its exchange

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Is the received data highly reliable and It can be relied upon and used

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Is the transmitted data the same as the one received

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Hardware and software problems hinder data exchange

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Do you think that the quality of the data affects the process of interaction between the two parties

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The data exchanged is clear and accurate

- Agree
- Strongly agree
- Disagree
- Strongly disagree

the equipment and software help to collect information and send it when needed

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Various data sources facilitate data exchange

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The method of building data storage and processing sites meets the needs of the data exchange process

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The devices and equipment used are flexible and easy to develop

- Agree
- Strongly agree

- Disagree
- Strongly disagree

Are there barriers to interoperability and data exchange

- Agree
- Strongly agree
- Disagree
- Strongly disagree

If there are obstacles, what kind of these obstacles are

- Technical (related to the technologies used within the organization).
- Institutional (related to laws and regulations within the institution).
- Conceptual (relates to the compatibility of concepts and meanings between different data sources).
- There are no obstacles

It is possible to exchange data even with obstacles

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The exchange of data requires high efforts and costs that affect the partnership

- Agree
- Strongly agree
- Disagree
- Strongly disagree

There is easiness in exchanging data with a third party

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Systems are conceptually compatible (same generic formulas, same information meanings)

- Agree
- Strongly agree
- Disagree
- Strongly disagree

Platforms at both ends of the data exchange are technically compatible

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The equipment and software perform their functions excellently

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The quality of the equipment and software positively affects performance and decision-making

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The responsibilities and powers are quite clear

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The communication protocols are compatible between the two parties

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The time of data exchange between the time it is requested and the time it is used positively affects the speed of decision-making

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The quality of the data exchanged meets the needs of the decision-makers

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The amount of the data that has been transmitted is the same as the data that has been used

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The transmitted data is identical to the recipient data

- Agree
- Strongly agree
- Disagree
- Strongly disagree

The cost of sharing data

- High
- Medium
- Low