UNIVERSITY OF NAIROBI
FACULTY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

AN ASSESSMENT ON THE ACCEPTANCE OF BIG DATA IN STATISTICS

ALLAN GATHURU WAIRIMU
P54/35452/2019

SUPERVISOR
CHRISTOPHER A. MOTURI

A project report submitted in partial fulfillment for the degree of Master of Science in Information Technology Management in the Department of Computer Science of the University of Nairobi.

JULY 2021

DECLARATION

This research project is my original work and has not been presented to any other university for the award of a degree.

Signature: Date: **31ST JULY 2021**

Wairimu Allan Gathuru

P54/35452/2019

This research has been submitted for examination with my approval as the University Supervisor.

Signature: Date: **August 2, 2021**

Christopher A. Moturi

Faculty of Science and Technology

Department of Computer Science

DEDICATION

This research is dedicated to my Grandmother who though having no formal education valued it and encouraged me to attain it.

ACKNOWLEDGEMENT

I thank God for enabling me to reach this milestone, His provision, care, favor, and Grace. I highly acknowledge my supervisor, Mr. Christopher Moturi for his guidance and his commitment. He is as good as they come. I thank Prof. Daniel Orwa for his insight and contribution to shape this study. I thank my colleagues Eric and Edith for their encouragement, critique, and contributions. I highly acknowledge Dr. Mary for her review and advice. Finally, I thank my family – Mum, Kabari, Njoki, Wangui, Wairimu, and Kathugu for their love, support, and encouragement.

ABSTRACT

The data revolution which has caused an explosion of data volumes and increased data demands is expected to have a big impact in research institutes. To support this data revolution and improve the quantity, frequency, disaggregation, and availability of relevant statistics, there is an essence to use Big Data in statistics. The study sought to assess the adoption of Big Data in research institutes in Kenya, establish the risks and challenges of using Big Data in statistics, identify the determinants of adoption of Big Data in statistics, and validating the relevance of TAM based model for predicting the adoption. Big Data is a transformative tool for statistics and has great potential to fill data gaps, leveraged to reduce costs and improve the availability of data to monitor development goals. The study used a descriptive survey where quantitative data was collected using self-administered questionnaires. The data was collected from sampled staff sampled from research institutes. Data were statistically analyzed using Stata. Composite reliability was used to assess reliability while Factor loadings and average variance extracted were used to assess convergent validity. Descriptive statistics for each construct of the TAM-based model were generated. The test of the structural model which includes estimating the path coefficients was done using Structural equation Modelling. The study found that research institutes are adopting Big Data in statistics by developing and using Big Data strategies. Legal and regulatory issues; gaining access to data; gaining access to associated methodology and metadata; establishing dataset quality are the main challenges of using Big Data in statistics. Inconsistent access and continuity; privacy breaches and data security; resource constraints and cut-backs; and resistance of Big Data providers and populace were noted as the most prominent risks. The study establishes that external influence, subjective norms, perceived usefulness, compatibility, attitude towards use, and self-efficacy as the key factors influencing acceptance of Big Data in Statistics. The limitation of the study was that Market research companies, credit reference bureaus, private research institutes, and Big Data Analytics companies deal with statistics and were not included in the study. Research institutes agree that Big Data can complement traditional sources of data to generate statistics and are ready to adopt it. However, the risks and challenges highlighted in the study must be overcome for successful adoption. The study recommends sensitization, training, and capacity building of data professionals, resolving of legal and regulatory issues, improvement of statistical methodologies of sampling and analysis, and allocation of more resources to Big Data projects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

APA	American Psychological Association
AVE	Average Variance Extracted
BDA	Big Data Analytics
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CMIN	Chi-square value
COVID-19	Coronavirus Disease 2019
CR	Composite Reliability
DF	Degree of Freedom
DOI	Diffusion Of Innovation
FSO	Federal Statistical Office
GIS	Geographic Information Systems
GPS	Geographical Positioning System
GWG	Global Working Group
ICT	Information Communication Technology
KNBS	Kenya National Bureau of Statistics
NFI	Normed Fit Index
NSIs	National Statistical Institutes
NSO	National Statistical Office
NSS	National Statistical System
ODK	Open Data Kit
OECD	Organisation for Economic Co-operation and Development
PARIS21	The Partnership in Statistics for Development in the 21st Century
SDGs	Sustainable Development Goals
SEM	Structural Equation Modeling
TAM	Technology Acceptance Model
TOE	Technology, Organization, and Environment Model
TPB	Technology and Planned behavior
TRA	Theory of Reasoned Action
TTF	Task Technology Fit
UN	United Nations
UNSD	United Nations Statistics Division
UTAUT	Unified Theory of Acceptance and Use of Technology

CHAPTER 1: INTRODUCTION

1.1 Background of the study

The emergence of Big Data is poised to disrupt organizations that deal with the production and analysis of statistics (Struijs et al., 2014). National Statistical Institutes (NSIs) are required to publish official statistics generated by the National Statistical System (NSS) for their respective countries all over the world (Kitchin, 2015). All organizations inside a country that manage, create, or publish statistics on behalf of their governments are members of the NSS (OECD, 2002). As a norm, NSS organizations usually use data collected from national surveys and censuses to generate statistics. However, using administrative and Big Data sources to generate official statistics can be beneficial (Mohd Din et al., 2017).

The United Nations Statistical Commission(UNSC) which was founded in 1947, is the topmost statistical organization(*UNSD*, 2021). It brings together heads of country statistical institutes. It makes the highest decisions, sets standards, and develops concepts and methods for international statistical activities (*UNSD*, 2021). The United Nations Statistical Commission's role is to promote the development of standard national statistics indices; synchronize statistical work; develop the central statistical services of the secretariat; and guide UN organs on general issues on statistics (UNSD, 2021).

During the 45th meeting of the UN Statistical Commission in 2014, the Global Working Group(GWG) on Big Data for Official Statistics was created (Economic & Council, 2016). The UN GWG provides strategic vision, direction, and coordination of a global programme on Big Data for official statistics, including for indicators of the 2030 agenda for sustainable development. The UN GWG promotes the training of personnel, sharing of Big Data experience across the data practitioners fosters. It also fosters advocacy to sensitize people and build trust in Big Data (Economic & Council, 2016). It addresses facilitating conditions to promote the adoption of Big Data(*About — UN-CEBD*, 2021).

Over the years the UN GWG has been organizing global conferences the inaugural one in Beijing, China 2014, the second one was in Abu Dhabi, United Arab Emirates 2015, the third one, Dublin, Ireland 2016, the fourth one in Bogota, Colombia 2017 where the **Bogota Declaration** was coined, the fifth one in Kigali, Rwanda 2019 where the **Kigali Declaration** was coined and the last one organized online in collaboration Federal Statistical Office (FSO) of Switzerland, 2020 (*Events — UN-CEBD*, 2021). The Bogota and the Kigali declarations seek to promote the adoption of Big Data to generate statistics.

The United Nations, the European Commission, the Organisation for Economic Co-operation and Development (OECD), the International Monetary Fund, and the World Bank formed the Partnership in Statistics for Development in the Twenty-First Century (PARIS21) in 1999. (About PARIS21, 2021). Its major purpose is to help low- and middle-income nations accomplish national and international development goals and decrease poverty by promoting statistical capacity development, advocating for the use of accurate data in decision-making, and coordinating donor assistance for statistics (About PARIS21, 2021).PARIS21 encourages the improved use and development of statistics in developing countries. It has effectively built a global network of statisticians, policymakers, analysts, and development practitioners who are devoted to making decisions based on evidence (About PARIS21, 2021).

At the continental level, the Pan-African Institute for Statistics forms part of the process of implementing best practices for the Statistical institutes in Africa and a realization of the partnership between the European Union and the African Union (African Union, 2021).

The Kenya National Bureau of Statistics (KNBS) was formed in 2006 by the Statistics Act, which mandated it to be the primary government agency for handling statistical data and information ('KNBS Mandate', 2020). Official country statistics is essential since it provides the populace with data on all situations of their lives whether demographic, social, or environmental (UNECE, 2014). The Bureau compiles and disseminates statistics to citizens to honor their right to information. The bureau collects data from surveys and censuses and some of its products include Kenya Population and housing census, Consumer Price Index, Lead Economic Indicators, Quarterly Gross Domestic Product, etc. These surveys help the country track its development agendas like Medium Term Plan 3 (Big Four Agenda) and internationally agreed initiatives like Sustainable Development Goals (SDGs). Research institutions are part of the NSS as highlighted in the Kenya Strategy for Development of Statistics(KSDS) 2019-2023(KNBS, 2020).

1.2 Statement of the problem

Over the years official government statistics have been generated using the data collected from surveys, censuses, and administrative data These data sources are costly; frequency of collection is low hence rapidly become out-of-date; and usually with no spatial distribution since they are reported at the national level (Fritz et al., 2019). These data sources are not enough to measure all the country, regional and global targets like the UN SDGs. Advances in technology have led to a data revolution characterized by increasing data volumes and extensive data demands (UN, 2013). To support this data revolution, reduce data collection costs and improve volume, occurrence, spatial distribution, and availability of relevant statistics, there is a need to make use of Big Data to generate statistics (OECD, 2013). Big Data has great potential to fill data gaps and is a transformative tool for statistics (Landefeld, 2014). The use of Big Data can reduce data collection costs and increase the volume of data to report more indicators(OECD, 2013). This use of Big Data comes with various opportunities, challenges, and risks that will disrupt the production of Government statistics (Kitchin, 2015). The statistical system will indisputably feel the pressure from the alternative sources of data and hence need to re-evaluate the usability of those sources for the generation of evidence for decision making. As Big Data gains traction and attracts more data users, there is a need to forge a partnership and working methodologies with statistics or else statistics risk outmodedness(OECD, 2013). Moreover, without guidance and coordination, Big Data may be unusable in statistics and just add to the discordance of data challenges (OECD, 2013). This study sought to assess the adoption of Big Data in research institutes in Kenya, establish the risks and challenges of using Big Data in statistics, identify the determinants of adoption of Big Data in statistics and validate the relevance of TAM-based model for predicting the adoption.

1.3 Purpose of the study

This research sought to assess the adoption of Big Data in research institutes, establish the risks and challenges of using Big Data in statistics, identify the determinants of adoption of Big Data in statistics, and validating the relevance of the TAM-based for predicting the adoption.

1.4 Objectives of the study

1. Determine how Big Data is being used in research institutes.
2. Establish the risks and challenges of using Big Data by research institutes.
3. Identify determinants of adopting Big Data among research institutes for statistical analysis.
4. Validate the research model using Structural Equation Modeling (SEM).

1.5 Research Questions

1. How is Big Data being used in research institutes?
2. What are the risks posed by the use of Big Data by research institutes?
3. What are the challenges affecting the use of Big Data by research institutes?
4. What determines the use of Big Data in research institutes?
5. To what extent do the dimensions of the research model consistent with the adoption of Big Data in the research institutes?

1.6 Scope of the Study

The attention of the research will be Government and Non-Government research institutes located in Kenya.

1.7 Significance of the study

The study will guide research institutes on the key determinants to prioritize when adopting Big Data in statistics. This study will highlight the challenges and risks of the use of Big Data in statistics which will guide research institutes as they adopt Big Data in their operations. Policy-makers will be guided on opportunities of Big Data to track real-time indicators.

1.8 Assumptions and Limitations of the Study

Market research companies, credit reference bureaus, private research institutes, and Big Data Analytics companies deal with statistics and were not included in the study.

1.9 Definition of operational terms

National Statistical System - A combination of entities within a nation that deals with statistics on behalf of the government (*NSS*, 2017).

Big Data – It is a field that deals with data that are too large or complex ('Big Data', 2020).

Official statistics – This is statistics generated and published by the government, public or international entities as a public good ('Official Statistics', 2020).

Exhaust Data – This is additional data collected purposely or inadvertently from digital transactions without an initial or specific purpose for its collection (OLeary & Storey, 2020).

Structural Equation Modelling– This is a procedure to evaluate models and it consists of various methods such as paths, confirmatory factor, structural relation, and covariance structure analysis (Hair et al., 2006).

CHAPTER 2: LITERATURE REVIEW

This chapter presents material gathered from previous literature that is related to this study and focuses mainly on the concepts that were canvassed and investigated in evaluating the adoption of Big Data in statistics.

2.1 Adoption of Technology in Organizations

Organizations are adopting technology to promote service delivery, increase transparency, save costs and improve efficiency. According to (Consoli, 2012) the inhibitor factors that discourage investment and adoption of ICT are categorized into financial, infrastructural, organizational, and technological factors. Consoli (2012) further categorizes the determinant factors into individual, organizational, environmental, technological, and economic factors and benefits into performance, growth, expansion, and new products. National Statistical Institutes around the world have adopted technology in data collection where Internet response option; telephone interviewing; and hand-held devices have been considered as recommended by the United Nations, 2015 in its report on recommendations for the 2020 censuses of population and housing. Various census management software has been adopted for data management and data processing. Geographic Information Systems (GIS) have been adopted as a tool to support the process of conducting data collection. During the dissemination of outputs, digital maps are increasingly playing an important role.

2.2 The 2019 Paperless Census

According to KNBS (2019), Kenya adopted the use of mobile technology to collect and transmit data during the 2019 census as recommended by the United Nations for the 2020 round of censuses. For the first time, mobile technology was used in the capture and transmission of cartographic mapping and enumeration with the mobile devices used for data collection assembled locally (KNBS, 2019). Smartphones and tablets embedded with Geographical Positioning System (GPS) were used to pick coordinates of homesteads, households, and other points of interest. Satellite images and aerial photographs were used to prepare maps for both rural and urban areas. Field mapping was done using Open Data Kit (ODK) while ArcGIS was used for digitization and map production.

2.3 The Data Revolution

Twenty million individuals in Kenya from 3 years and above own a mobile phone, use the internet, and use a computer while 4% from fifteen years and above searched and bought goods and services online (KNBS, 2019). This usage of ICT is driving the data revolution in Kenya which is being characterized by increased data volumes and high data demands (UN, 2013). For this data to be meaningful to the populace, it must bridge the national data disparities. Better data and statistics will help governments evaluate progress, make evidence-based choices, and increase accountability, according to the United Nations High-Level Panel on the global development framework for the Millennium Development Goals (MDGs) in their 2019 report (UN, 2013).

In developing countries, National Statistical Offices will play a central role in the period of data revolution(Badiee et al., 2017). This is because of their expertise, their quality assurance role, have established legal and regulatory frameworks, and are already mandated to handle data as highlighted in the PARIS-21 road map (PARIS21, 2015). The roadmap identifies technology and innovation as one of its pillars and categorizes them into action areas (PARIS21, 2015). Big Data has the potential to reduce costs and improve the accuracy of government statistics and this is acknowledged by statistics practitioners (Landefeld, 2014).

2.4 Characteristics and Sources of Big Data

Big Data can be categorized using 5 Vs of Volume(amount of information), Velocity(speed of generation), Variety(different types), Veracity(quality and accuracy), and Vulnerability(risk to privacy and confidentiality) (Tam & Halderen, 2020). It can also be categorized using the 3C's of Crumbs, Capacities, and Community(Letouzé & Jütting, 2015). People are very dependent on technology to handle their needs and our online transactions in social media, e-commerce, online searches, reading habits, blogs visited, online movies and music, and travel leaves a lot of data and digital footprints and can be a source of Big Data(MacFeely, 2018). Sources of Big Data include administrative data (banking records, hospital records, tax records), transactional data (credit and debit cards purchases, mobile money), sensor networks data(satellite imaging, traffic sensors, climate sensors), tracking devices data(GPS, mobile data), Behavioral data(online data) , and opinion data sources, e.g. social media activity(Tam & Clarke, 2015).

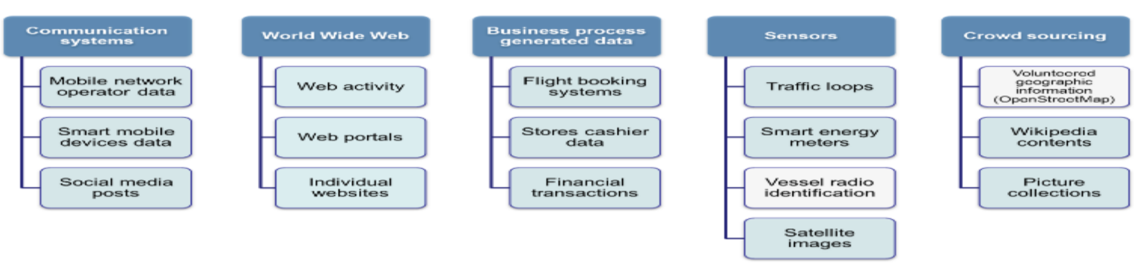

Figure 2.1 Classification of Big Data sources (Demunter, 2017)

2.5 Big Data in Official Statistics

The United Nations is driving the use of new technologies and data sources in the generation of statistics(UN Statistical Commission, 2017). The Bogota Declaration which seeks to promote and facilitate data sharing by proposing platforms and partnerships was coined and adopted by the United Nations Global Working Group in November 2017 (UNGWG on Big Data for Official Statistics, 2017).

Our lives are highly ICT dependent and every digital transaction and activity executed leaves behind trails of Big Data (MacFeely, 2018). These trails can be used by data practitioners to generate new statistics or complement existing ones and with the advance in technology, these data can revolutionize statistics (MacFeely, 2019). The statistical and governance challenges presented by Big Data can be cauterized into legal, ethical, technical, reputational, and expectation management (Kitchin, 2015). Big Data is not necessarily easily accessible and does not guarantee cost savings in the generation of statistics(MacFeely, 2019).

The data generated from surveys, censuses, and administrative data can be regarded as small data and has some characteristics of Big Data (Kitchin, 2015). For example, the census may be exhaustive but the speed of generation is slow since it is carried out after ten years in Kenya, no variety since it has a limited number of structured questions, and has no flexibility since it cannot be altered in the middle of the exercise. On the other hand, Big Data generated from telecoms, supermarkets, and retail shops, traffic sensors and social media have the 5Vs and 3Cs of Big Data.

Across the world, as highlighted in Figure 2.2, statistical institutes worldwide have initiated Big Data projects. They are using exhaust data, digital content, and sensing data from web scraping, Google maps, call detail records, satellite, and Twitter. Some of the projects noted include infection prevention and control, developing water accounts, tourism monitoring, subjective wellbeing, movements across borders, and complementing the national agriculture census (Sangokoya et al. 2015).

Selected Country	Big Data Used	Data used	Projects in progress
Argentina	Exhaust data	Web scraping	Infection Prevention and Control Online
Brazil	Digital content	Google Maps Call Detail Records	Developing Water Accounts Tourism Monitoring
Colombia	Exhaust data	Web scraping	Infection Prevention and Control Online SIPSA
	Digital content	Call Detail Records	Monitoring Crime Activities CeluPop
	Sensing data	Satellite	Complementing the National Agriculture Census
Ecuador	Exhaust data	Web scraping	Infection Prevention and Control Online
	Digital content	Twitter Call Detail Records	Measuring Subjective Wellbeing Daytime Migration
Guatemala	Digital content	Call Detail Records	Monitoring Poverty Levels
Mexico	Digital content	Twitter	Subjective Wellbeing Subjective Wellbeing of Women Tourism Monitoring Movements Across Borders

Figure 2.2: Overview of Big Data projects in selected National Statistical Offices (Sangokoya et al., 2015)

2.5.1 Opportunities of Big Data for official statistics

Big Data possess great opportunities for statistics if data access issues and other challenges are resolved (MacFeely, 2018). It can entirely substitute, partially substitute, complement, improve estimation, or provide entirely new indicators that never existed before(Florescu et al., 2014). It can also contribute to the creation of a selection frame, connecting to other data, contribute to missing datasets, data confirmation, and data editing (Tam & Clarke, 2015).

2.5.2 Challenges of Big Data for official statistics

Scannapieco et al. (2013) envision dimension, quality, time dependence, and accessibility as the challenges of using Big Data in official statistics. Dimension affects the storage of data since Big Data has high volumes and requires advanced storage technologies like Hadoop and NoSQL databases like HBase, BigTable, and MongoDB and also processing since Big Data has high volume and veracity hence analyzing is a hard task (Scannapieco et al., 2013).

Gaining access to the required Big Data for assessment, experimenting, trialing and adoption is a challenge (Tam & Clarke, 2015). Although some Big Data is produced by public agencies, much Big Data is presently generated by private companies such as mobile phone, social media, utility, financial and retail companies and are valuable commodities to these companies, either providing a resource that generates competitive advantage or constituting a key product and is generally not publicly available for official or public analysis in raw or derived forms (Kitchin, 2015). Producers of Big Data gain their competitive advantage by maintaining their data locked and hence the accessibility challenge (Scannapieco et al., 2013). The use of probabilistic sampling in traditional statistics provides a theoretical framework that ensures a clear methodology and confidence in the figures based on sampling errors. Sivarajah et al. (2017) explain that other frameworks should be developed because most of the Big Data available cannot be adapted to this existing theoretical framework. This weakness is a pertinent issue and efforts should be focused on it.

2.5.3 Risks of Big Data for official statistics

The key risks relate to mission drift, reputation and trust, privacy and data security, access and continuity, fragmentation across jurisdictions, resource constraints and cut-backs, and privatization and competition(Kitchin, 2015).

Traditionally data have been generated to answer a specific set of queries but in the era of Big Data, this will be reversed with the wealth and cost-benefit of Big Data setting the agenda for what is to be measured hence official statistics may drift towards following the data, rather than the data being produced for the compilation of official statistics (Kitchin, 2015).

Opportunities, challenges and risks of big data for official statistics

Opportunities	Challenges	Risks
– Complement, replace, improve, and add to existing datasets – Produce more timely outputs – Compensate for survey fatigue of citizens and companies – Complement and extend micro-level and small area analysis – Improve quality and ground truthing – Refine existing statistical composition – Easier cross-jurisdictional comparisons – Better linking to other datasets – New data analytics producing new and better insights – Reduced costs – Optimization of working practices and efficiency gains in production – Redeployment of staff to higher value tasks – Greater collaboration with computational social science, data science, and data industries – Greater visibility and use of official statistics	– Forming strategic alliances with big data producers – Gaining access to data, procurement and licensing – Gaining access to associated methodology and metadata – Establishing provenance and lineage of datasets – Legal and regulatory issues, including intellectual property – Establishing suitability for purpose – Establishing dataset quality with respect to veracity (accuracy, fidelity), uncertainty, error, bias, reliability, and calibration – Technological feasibility – Methodological feasibility – Experimenting and trialing big analytic – Institutional change management – Ensuring inter-jurisdictional collaboration and common standards	– Mission drift – Damage to reputation and losing public trust – Privacy breaches and data security – Inconsistent access and continuity – Resistance of big data providers and populace – Fragmentation of approaches across jurisdictions – Resource constraints and cut-backs – Privatisation and competition

Figure 2.3: Risks, Challenges, and Opportunities of Big Data for official statistics (Kitchin, 2015)

2.6 ICT Adoption Models

The commonly used adoption models can be categorized into adoption at the organization level(Diffusion Of Innovation(DOI) and Technology, Organization and Environment(TOE)) or adoption at the individual level (Technology Acceptance Model(TAM), Technology and Planned behavior(TPB), Unified Theory of Acceptance and Use of Technology(UTAUT))(Oliveira & Martins, 2011).

2.6.1 The Technology–Organization–Environment (TOE) Framework

This theory explains that the technological, organizational and environmental contexts of an organization influence its possibility to adopt a particular technology (Baker, 2011).

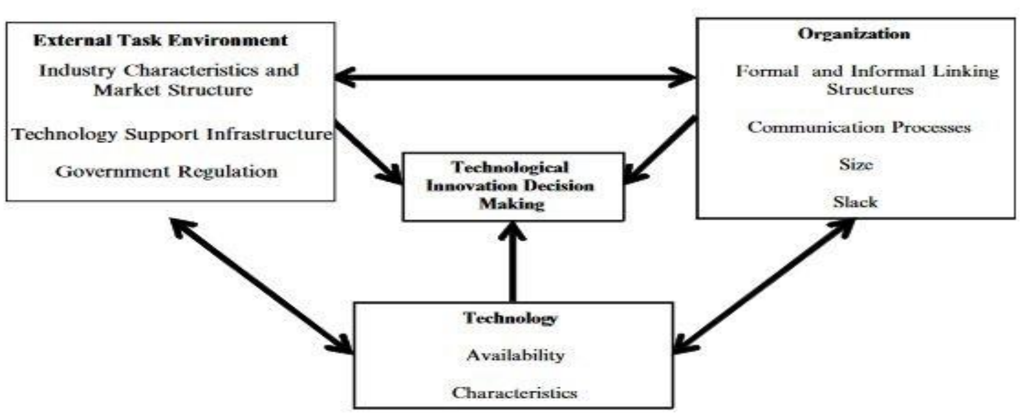

Figure 2.4: The technology–organization–environment framework Source: (Baker, 2011)

2.6.2 The Unified Theory of Acceptance and Use of Technology (UTAUT)

UTAUT recognizes social influence, facilitating conditions, performance expectancy, and effort expectancy as the key factors and experience, age, gender, and voluntariness as moderators influencing adoption of technology by an individual (Venkatesh et al., 2016)

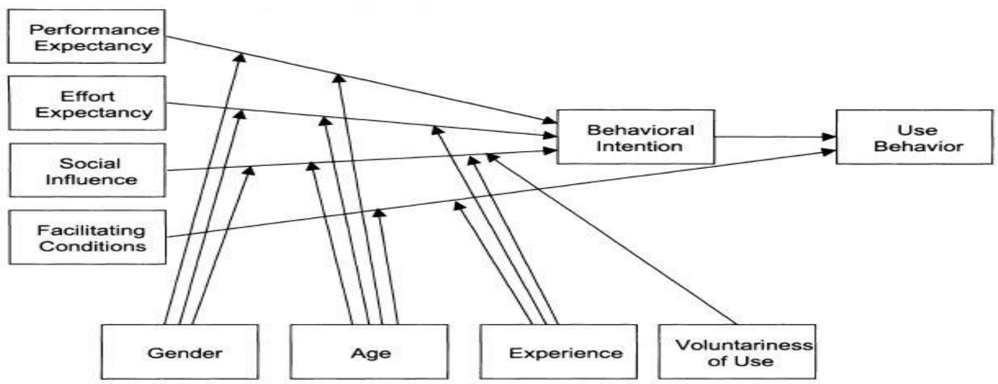

Figure 2.5: Unified Theory of Acceptance and Use of Technology Source: (Venkatesh et al., 2016)

2.6.3 Technology and Planned behavior (TPB)

This theory identifies attitudes towards use, subjective norms, perceived behavioral control as predictors, and intention to use a particular technology as a mediator between them and actual use of technology(White Baker et al., 2007).

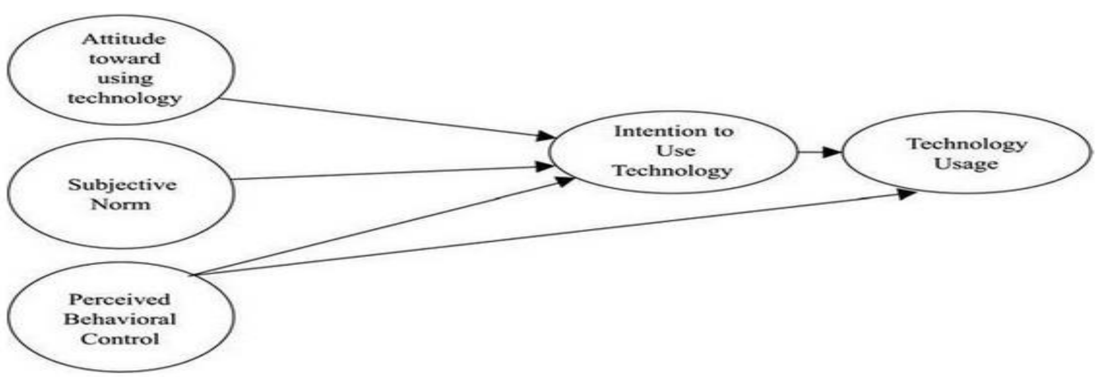

Figure 2.6: Theory of planned behavior (TPB) (technology-specific) Source: (White Baker et al., 2007)

2.6.4 Technology Acceptance Model (TAM)

This theory was developed by Davis (1989) building on Ajzen & Fishbein's (1980) Theory of Reasoned Action (TRA). It identifies perceived usefulness and perceived ease of use as determinants of an individual's likelihood to use technology. The two determinants are influenced by external factors that may be either social (language, skills, and facilitating conditions), cultural, or political (Beselga & Alturas, 2019).

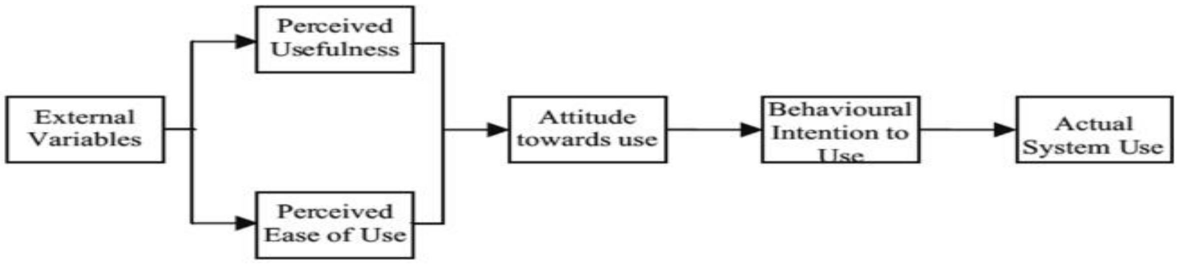

Figure 2.7: Technology Acceptance Model (TAM) Source: (Davis, 1989)

2.7 Conceptual framework

The study will adopt a model based on TAM developed by (Mohd Suki & Ramayah, 2010) and use it to assess the adoption of Big Data statistics. In this TAM-based model, the TAM is extended using the following constructs: facilitating conditions, peer influence, self-efficacy, external influence, compatibility, perceived behavioral control, and subjective norms. From the model adopted, the following are some of the determinants of the adoption of technology.

2.7.1 Perceived Usefulness

This is the extent to which an individual rates benefits of a technology relating to the tasks being executed and hence accepts to use it (Davis, 1989). Table 2.3 shows different measurement constructs for Perceived Usefulness.

Table 2.1: Typical perceived usefulness constructs measurements

Measurement Constructs	References from literature
Make work faster, Increase work performance, Increase output, Efficiency, Makes work easier and work useful	Davis, 1989

2.7.2 Perceived Ease of Use

This is the level to which uses believe that accepting and using technology for instance Big Data is easy and will not require a lot of learning and effort (Davis, 1989). Table 2.4 highlights the different constructs that can be used to measure PEOU.

Table 2.2: Typical Perceived ease of use constructs measurements

Measurement Constructs	References from literature
Easy to learn and use, easy to comprehend and become skillful, Controllable, Flexible	Davis, 1989

2.7.3 Compatibility

It is an important construct of the diffusion of innovation concept. It is more probable for people to accept a technology that is compatible with their work (Tornatzky & Klein, 1982).

Table 2.3: Typical Compatibility constructs measurements

Measurement Constructs	References from literature
Fit well with work, Fit in work style, Compatible with work	Mohd Suki & Ramayah, 2010

2.7.4 Social Influences

This determinant is derived from Venkatesh et al.'s (2016) UTAUT. Social influence means that in some cases individuals may use technology to comply and please others especially people with authority but not necessarily out of their feelings and beliefs concerning the technology (Davis et al., 1989).

Table 2.4: Typical Social Influences constructs measurements

Measurement Constructs	References from literature
Interpersonal Influence: peer influence, peer opinion	Venkatesh et al., 2016,
External influence: mass media influence, popular press influence	Mohd Suki & Ramayah, 2010

2.7.5 Facilitating Conditions

This determinant is derived from Venkatesh et al.'s (2016) UTAUT. These conditions that may affect the adoption of technology include resource factors; technology factors; capacity building and user support; and external controls such as policies, regulations, and legal environment.

Table 2.5: Typical Facilitating Conditions constructs measurements

Measurement Constructs	References from literature
Resources, knowledge, and ability are accessible	Venkatesh et al., 2016, Mohd Suki & Ramayah, 2010

2.7.6 Self-Efficacy

This is the belief by an individual that they can be able to use a particular tool or technology to execute a particular task (Bandura, 1986).

Table 2.6: Typical Self-Efficacy constructs measurements

Measurement Constructs	References from literature
Comfortable using, use reasonably well, user assistance	Ajzen, 1991, Mohd Suki & Ramayah, 2010

2.7.7 Subjective Norms

This is a people's influence and control over an individual’s beliefs on whether they should use a tool or a technology to perform a particular task (Ajzen & Fishbein, 1980). It is determined by how an individual is conditioned by both peer influence and external influence.

Table 2.7: Typical Subjective Norms constructs measurements

Measurement Constructs	References from literature
User acceptance	Davis, 1989
Beliefs	Ajzen & Fishbein, 1980

2.7.8 Perceived Behavioural Control

It encompasses facilitating conditions and self-efficacy and reflects an individual’s belief regarding the availability of resources to use technology and their ability to use it (Ajzen, 1991).

Table 2.8: Typical Perceived Behavioural Control constructs measurements

Measurement Constructs	References from literature
availability of opportunities, control, availability of resources, knowledge, and ability	Ajzen, 1991, Mohd Suki & Ramayah, 2010

2.7.9 Attitude towards Use

It mediates between perceived usefulness & ease of use of a particular technology and the intention to use the technology(Davis, 1989). Table 2.11 shows different measurement constructs for Attitude towards use.

Table 2.9: Typical Attitude towards use constructs measurements

Measurement Constructs	References from literature
Good, Favorable, Positive influence, Valuable, Flexibility	Weng et al., 2018

2.7.10 Intention to use Technology

This is the possibility of an individual to use a technology(Davis, 1989). Table 2.12 shows different measurement constructs for intention to use technology.

Table 2.10: Typical Intention to use tool constructs measurements

Measurement Constructs	References from literature
Love Using, Intend to Use, Think it's more helpful	Weng et al., 2018

The research model is thus generated as shown in Figure 2.8 showing the hypotheses between the different constructs

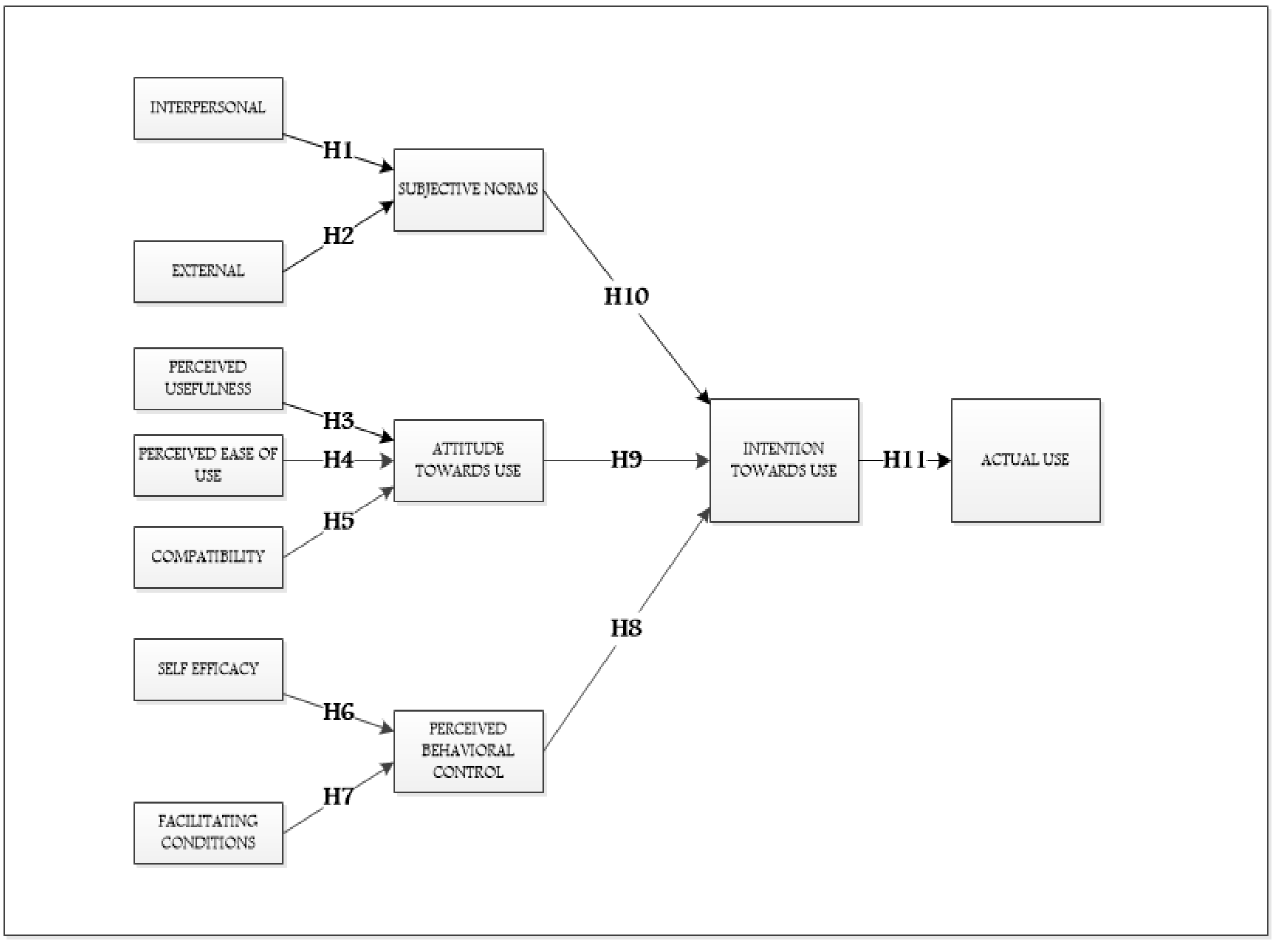

Figure 2.8: Research Model

Being a one-time cross-sectional study without a follow-up is constraining since it cannot provide a definitive relation between intent and actual behavior relation. Therefore, H11 will be dropped.

The hypothesis for the research framework is formulated as follows

Table 2.11: Framed Hypothesis

H1	Interpersonal influence has a significant effect on Subjective norms.
H2	External influence has a significant effect on Subjective norms.
H3	Perceived usefulness has a significant effect on Attitude towards the use of Big Data in statistics.
H4	Perceived ease of use has a significant effect on Attitude towards the use of Big Data in statistics.
H5	Compatibility has a significant effect on Attitude towards the use of Big Data in statistics.
H6	Self-efficacy has a significant effect on Perceived Behavior Control.
H7	Facilitating conditions have a significant effect on Perceived Behaviour Control.
H8	Perceived behavioral control has a significant effect on the Intention to use Big Data in statistics.
H9	Attitude has a significant effect on the Intention to use Big Data in statistics.
H10	Subjective norms have a significant effect on the Intention to use Big Data in statistics.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter contains the research design, the method used in collecting data, the data analysis methods, the data validity testing, and the ethical considerations of the study.

3.1 Research Design

To manage preconceived notions in research and safeguard reliability and validity, a research strategy must be designed (Al-Raqadi et al., 2015). The research study used a deductive approach since it was anchored on an already existing theory to conclude. This was achieved by the use of a descriptive survey where quantitative data was collected, cleaned, studied, and conclusions were drawn using the TAM-based model adopted and from a positivist perspective (Saunders et al., 2000).

3.2 Population

A population is the universe of characters with some similar recognizable features (Kothari, 2004). This study targeted professionals dealing with data and the generation of statistics in research institutes in Kenya. The respondents of the study were statisticians, Data Managers, Data Scientists, and data analysts, and statistical programmers.

3.3 Data Collection

Empirical research is widely done through the usage of surveys (Oates, 2005), hence the quantitative method is applied. A survey was carried out to collect data whereby self-administered questionnaires were sent to sampled respondents via email due to the existing COVID-19 pandemic. Where the questionnaire failed to work, telephone-based interviews were conducted.

The questionnaire was categorized into 11 segments where the first segment contained demographic attributes of the respondent whereas 10 segments sought to answer the dimensions of the TAM-based model used. Specific questions for each dimension were asked and answers were selected from a five-point Likert scale with a range: strongly agree =5; agree = 4; neutral = 3; disagree = 2; and strongly disagree=1.

3.4 Validity and Reliability

The research instrument was validated to reduce measurement errors (Kimberlin & Winterstein, 2008). Composite Reliability(CR), Average Variance Extracted(AVE) and factor loadings were employed to measure reliability and convergent validity (Hair et al., 2006).

3.5 Analysis of Data

Data cleaning was done to remove any errors and inconsistencies then coded before being imported to Stata software for analysis. Descriptive statistics analysis measures were carried out which included central propensities, distribution, and occurrence. Composite Reliability(CR), Average Variance Extracted(AVE) and factor loadings were employed to measure reliability and convergent validity (Hair et al., 2006). The path analysis method of structured equation modeling technique was used for validation of the conceptual model used. Pie charts, graphs, and tables were used to represent the data.

3.6 Ethical Consideration

The rights of the participants and responsibilities of the researcher as highlighted by (Oates, 2005) were fully adhered to. The participants involved were fully briefed about the purpose of the research in advance and responding to the study was optional. The anonymity of the participants was achieved by not collecting their names and locations. The research institution's identity was anonymized. The data collected was handled safely and securely to enhance confidentiality. The data collected was used for academic purposes only. Additionally, the participants were protected from incurring any cost when undertaking the survey. The privacy of the respondents was respected and all literature quoted was fully referenced.

CHAPTER 4: DATA ANALYSIS, RESULTS, AND DISCUSSION

This chapter contains the data analysis, presentation of the findings, and discussion of the results of the study. The demographic information of respondents, challenges, and risks of adoption of Big Data in Statistics and assessment of the proposed hypothesis against the results obtained is explored.

4.1 Demographic Characteristics

4.1.1 Response Rate

From the 96 respondents targeted 64 responses were received, which translated to a response level of 67% which is considered good (Mugenda & Mugenda, 2003).

4.1.2 Respondents Demographic Characteristics

The demographics of the respondents are highlighted in table 4.1.

Table 4.1: Respondents demographics

		Percentage
Gender	Male	61
	Female	39
Profession	Statisticians	46
	Data Analyst	25
	Data Scientist	18
	Data Manager	9
	Statistical Programmer	2
Age Bracket(Yrs.)	18-24	11
	25-34	48
	35 -44	27
	45 -60	14
	Over 60	0
Work Experience(Yrs.)	0 - 5	33
	6 -10	44
	11 - 20	11
	Over 20	12

4.2 Big Data Usage in research institutes, Risks, and Obstacles

4.2.1 Big Data Strategy

Forty-eight percent of the respondents confirmed that their institutes have a Big Data Strategy in place. The statistics show that half of the research institutes from whom the respondents were sampled have a Big Data Strategy which is a step in the right direction.

4.2.2 Big Data Usage in research institutes

Four questions to measure whether Big Data has been adopted at the organization level were asked and the results are summarized in figure 4.6. 91% of the respondent's institutes foresee a role of Big Data in the execution of their work. This shows that a majority of institutes prioritize Big Data and foresee its incorporation in their operations.

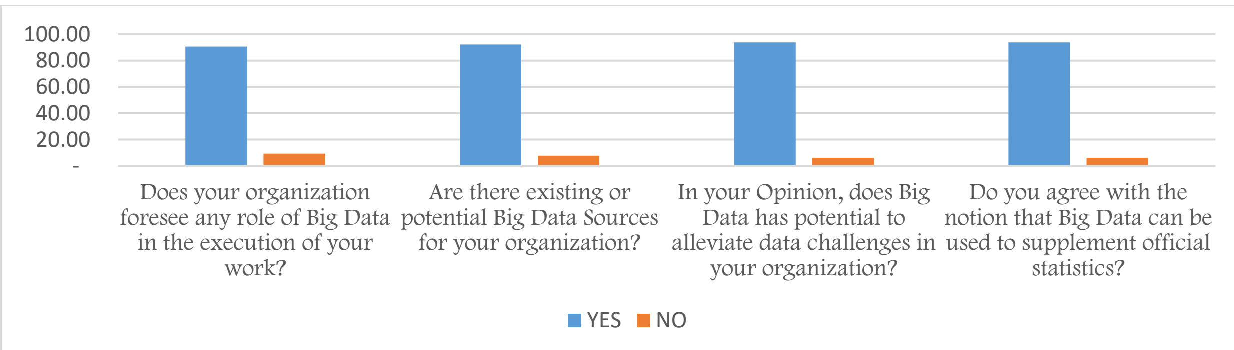

Figure 4.1: Big Data Usage in Research Institutes

92% of the respondents have existing potential Big Data sources that they can use in their work. This can be deduced to mean that there exists a lot of unutilized data that can be used in the analysis and generation of statistics.

94% of the respondents think that Big Data has can improve data challenges in their institutes. This positive response shows that the generation of statistics using Big Data will be adopted easily since the professionals dealing with data recognize its potential usefulness. 94% of the respondents agree with the notion that Big Data can supplement official statistics. This can be deduced to mean that professionals in the data sector welcome the rise of new data sources to complement the traditional sources of data.

4.2.3 Big Data Risks

Figure 4.7 details the sentiments of the respondents regarding the major obstacles highlighted. In concurrence to Kitchin (2015) observations, the most prominent risks as per the respondents of the survey were: inconsistent access and continuity reported by 69% of the respondents;

privacy breaches and data security reported by 66% of the respondents; resource constraints and cut-backs reported by 55% of the respondents; and resistance of Big Data providers and populace reported by 47% of the respondents. Privatization and competition 25%, damage to reputation and losing public trust 23%, and mission drift 23% were the least reported by the respondents.

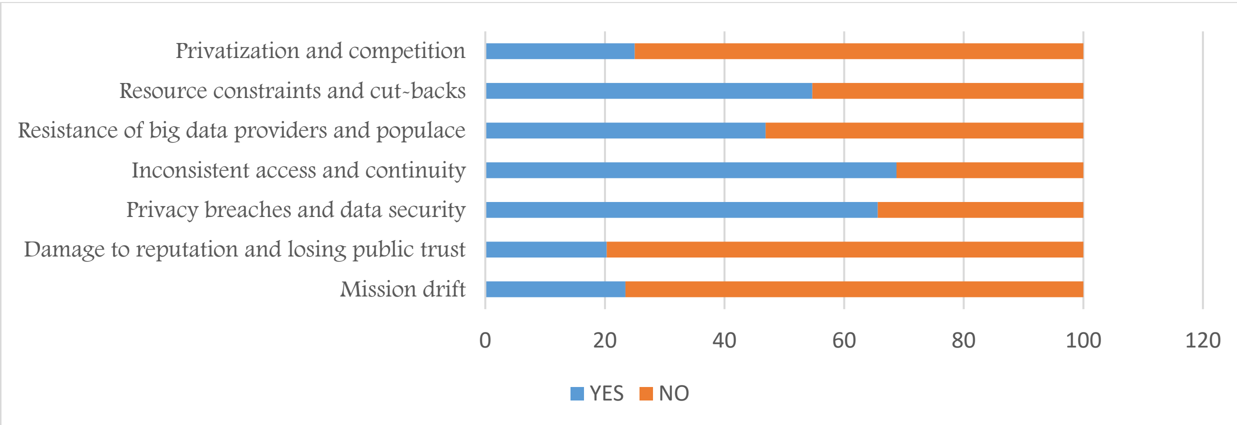

Figure 4.2: Major Big Data Risks

4.2.4 Big Data Obstacles

Figure 4.8 details the respondent selection to the question regarding the Big Data obstacles. These obstacles had been adopted from (Kitchin, 2015) and submitted to the respondents for validation. 70% of the respondents reported legal and regulatory issues as major obstacles. Fifty-six percent of the respondents reported gaining access to data as an obstacle which was proportional to gaining access to associated methodology and metadata. This signifies that data sharing is a major challenge. 55% of the respondents reported establishing dataset quality as a major obstacle. Since Big Data is not collected for statistical analysis, quality issues arise from the lack of standardization and representativeness. 39% of the respondents reported establishing suitability for purpose as a major obstacle. Institutional change management and guaranteeing inter-organizational collaboration and common standards were the least reported with 31% and 22 % selecting them respectively.

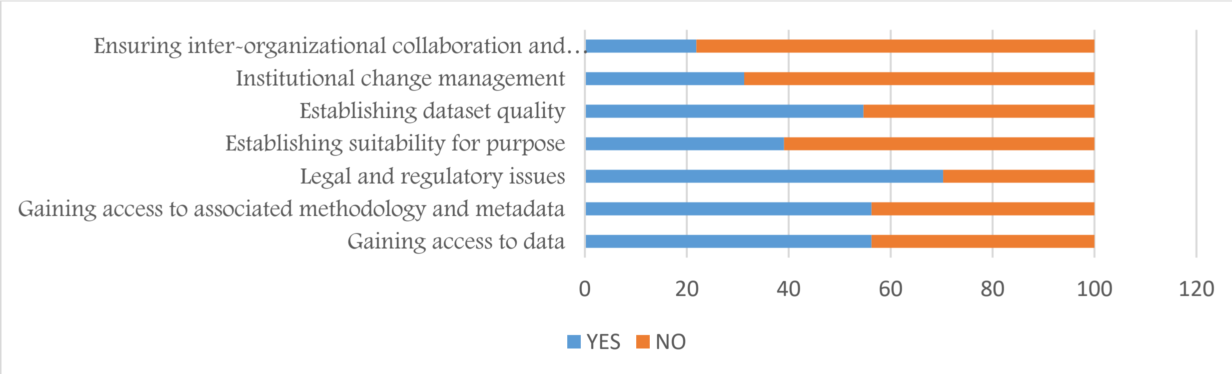

Figure 4.3: Major Big Data Obstacles

4.3 Model Descriptive Statistics

4.3.1 Interpersonal/Peer influence(IP)

Thirty-four percent of respondents strongly agreed on the first item. Some respondents between 18% and 28% were neutral on the measurement item. A low percentage ranging from 0% to 17% either disagreed or strongly disagreed.

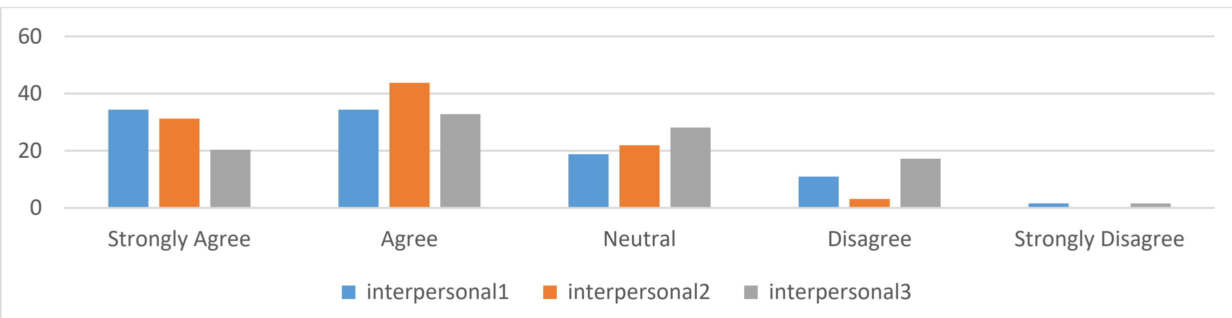

Figure 4.4: Percentage responses on Interpersonal/Peer influence

4.3.2 External Influence(EI)

48% of respondents strongly agreed on the first measurement item. A significant number ranging from 9% to 30% were neutral on the items given. Between 0% and 20% of the respondents either disagreed or strongly disagreed on the measurement items.

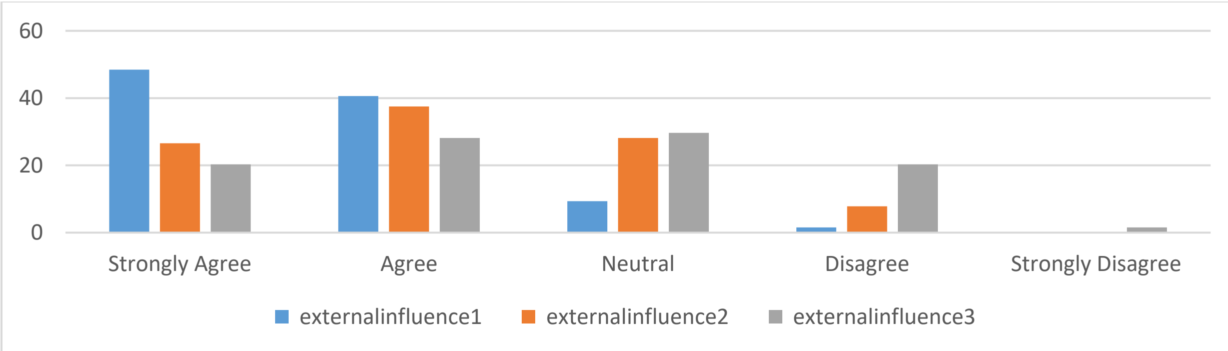

Figure 4.5: Percentage responses on External Influence

4.3.3 Perceived Usefulness(PU)

Between 28% and 40% strongly agreed on the four measurement items. Between 39% and 53% agreed on the four measurement items. A significant number were neutral with responses ranging from 17% to 22%. Very few respondents disagreed or strongly disagreed on the perceived usefulness of Big Data in statistics with the responses ranging from 0% to below 2%.

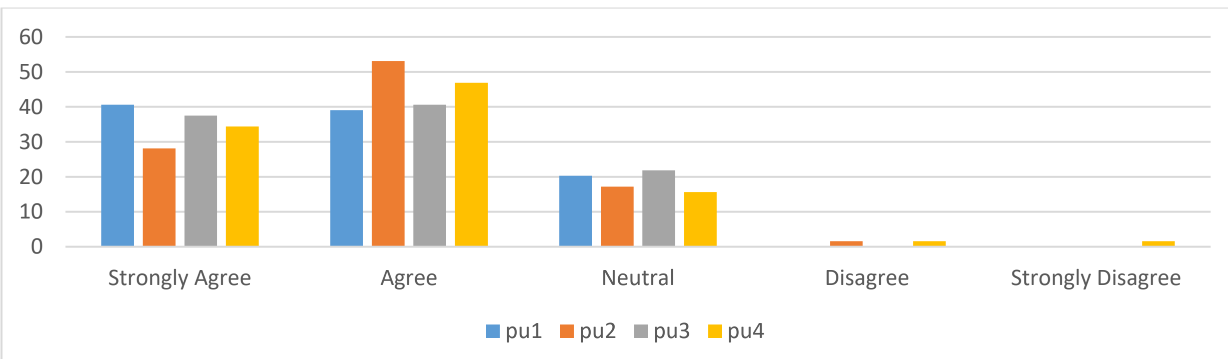

Figure 4.6: Percentage responses on Perceived Usefulness

4.3.4 Perceived Ease of Use(PEOU)

The results show that between 19% and 44% responded positively to the measurement items. Between 17% and 37% are neutral on the PEOU of Big Data in Statistics. Between 3% and 11% disagreed with the measurement items and an almost similar number of respondents below 2% disagreed on the four measurement items of perceived ease of use.

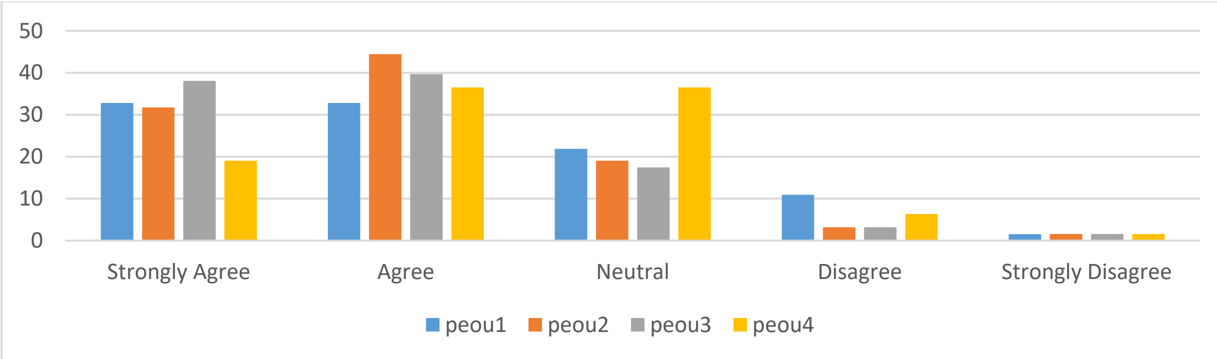

Figure 4.7: Percentage responses on Perceived Ease of Use

4.3.5 Compatibility(CP)

Over 25% of the respondents answered positively on the three Likert items with a high percentage of 46% agreeing on the second Likert item. Between 13% and 25% of the respondents were neutral on the measurement items of compatibility. A very low percentage ranging from 1% to 3% either agreed or disagreed.

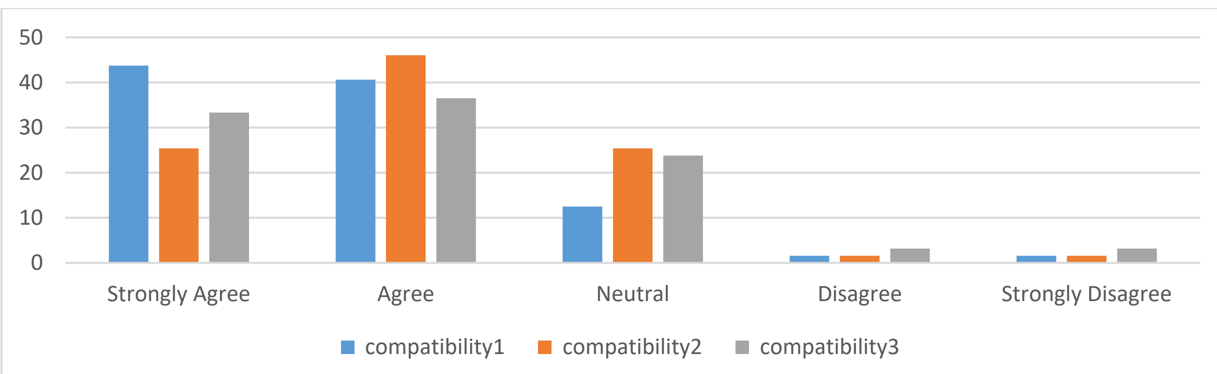

Figure 4.8: Percentage responses on Compatibility

4.3.6 Self-Efficacy(SE)

The measurement items for self-efficacy elicited mixed responses as shown in figure 4.14. The choice of strongly disagrees registered a few responses ranging from 0% to 5% on the three measurement items. Between 13% and 43% strongly agreed or agreed, were neutral, or disagreed on the measurement items.

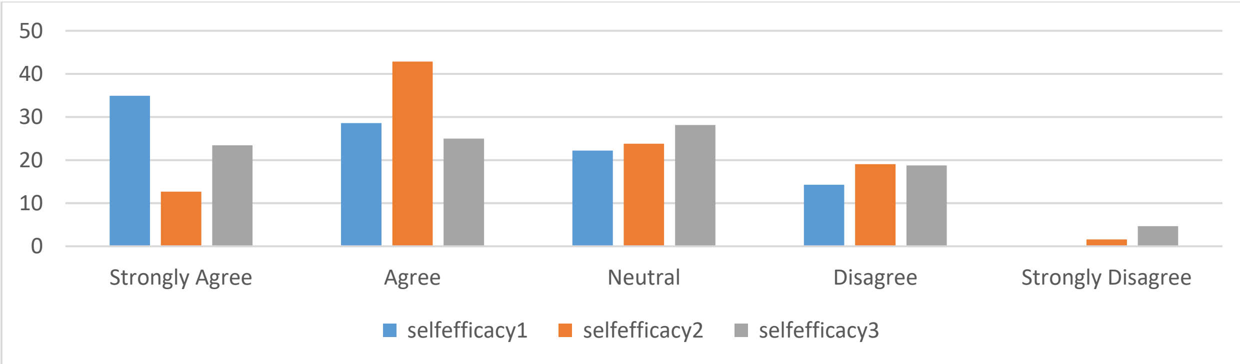

Figure 4.9: Percentage responses on Self-Efficacy

4.3.7 Facilitating Conditions(FC)

The responses for the three measurement items on facilitating conditions were spread across the five measurement categories as shown in figure 4.15. Strongly disagree got responses ranging from 3% to almost 10%. The first and second measurement items posted a 28% disagreeing which is a large proportion. Between 9% and 32% were either neutral, agreed, or disagreed with the measurement items.

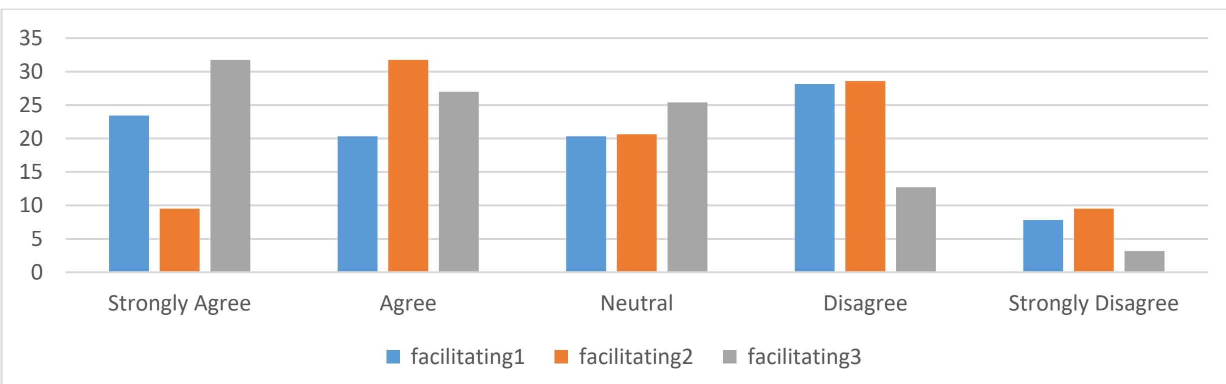

Figure 4.10: Percentage responses in facilitating conditions

4.3.8 Perceived Behavioral Control(PBC)

As detailed in figure 4.16, the majority of the respondents were either neutral or agreed with the measurement items of perceived behavioral control ranging from 22% to 41%. Between 0% and 30% either agreed or strongly disagreed with the measurement items.

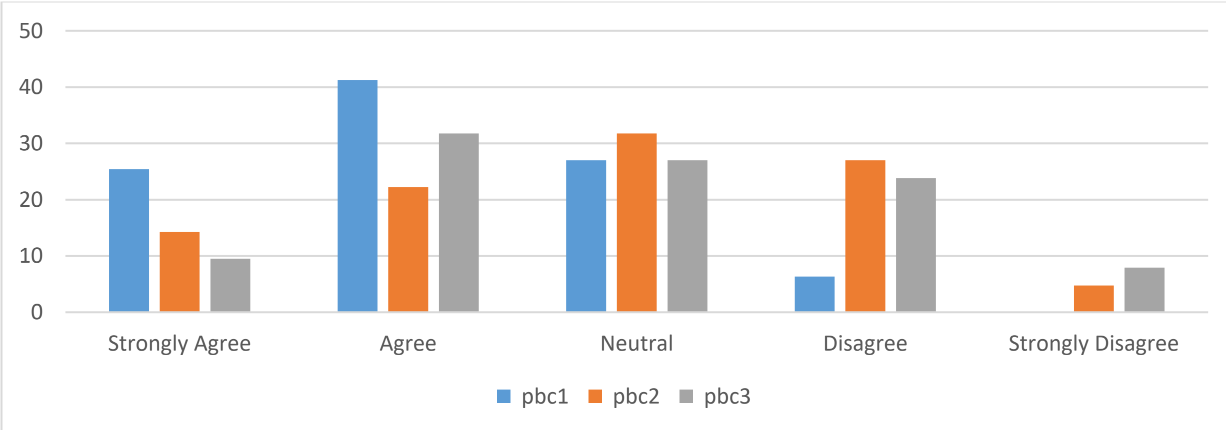

Figure 4.11: Percentage responses on Perceived Behavioral Control

4.3.9 Attitude towards Use(ATU)

The results show that between 38% and 49% of the respondents answered positively to the three measurement items of ATU. A uniform 11% were neutral on the three measurement items with strongly disagree posting 0% on the three measurement items.

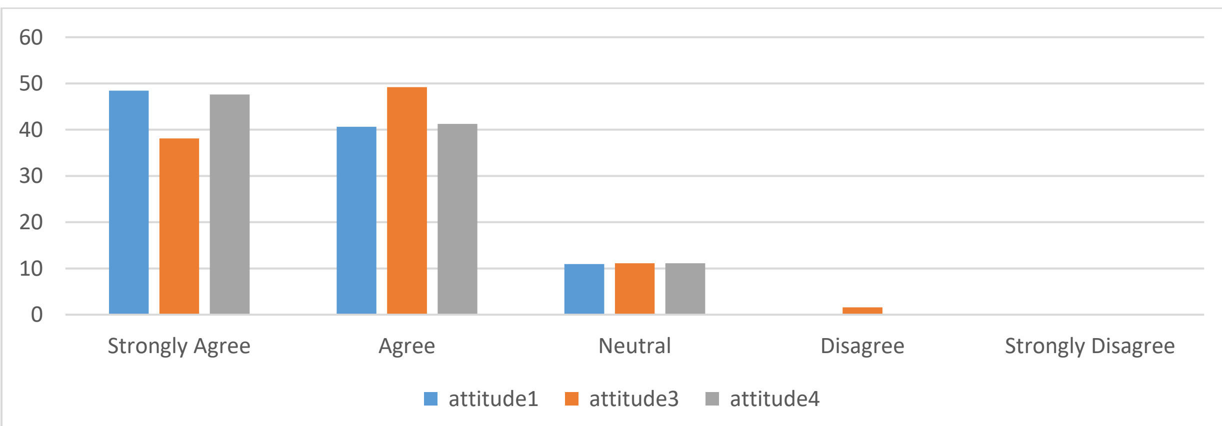

Figure 4.12: Percentage responses on Attitude towards use

4.3.10 Subjective Norms(SN)

Between 27% and 37% of the respondents were neutral, between 34% and 38% agreed, 23% to 27% strongly agreed, 6% to 10% disagreed while 0% strongly disagreed on the three measurement items of SN.

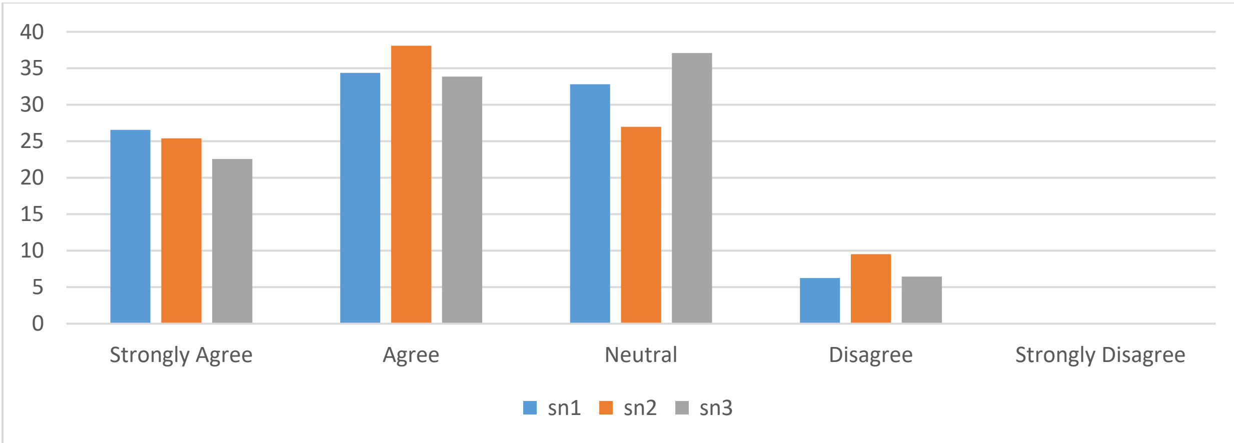

Figure 4.13: Percentage responses on Subjective Norms

4.3.11 Intention to Use(ITU)

No respondent disagreed with the three measurement items of intention to use. A low percentage ranging from 1% to 3% disagreed. Between 26% and 51% responded positively to the three measurement items of ITU.

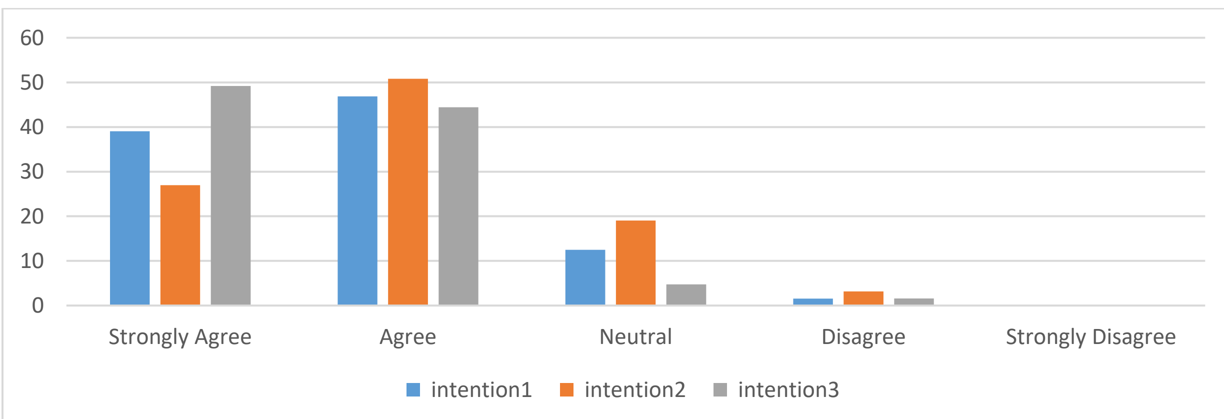

Figure 4.14: Percentage responses on Intention to Use

4.3.12 Summary of Model Descriptive Statistics

The means of all the constructs ranged from 3.11 to 4.36 indicating a general affirmative reaction to the constructs measured (Teo & Schalk, 2009). The standard deviations for all variables were less than one and this indicates that the item scores were relatively close to the mean scores. The kurtosis of constructs was between 2.0071 to 4.0649 and skewness was between -0.0898 to -0.8880. Skew indices should be below 3.0 and kurtosis indices should be below 8.0 hence the data were considered fairly normal with no severe problems (Kline, 2010).

Table 4.2: Summary of Model Descriptive Statistics

Construct	Mean	SD	Skewness	Kurtosis
Interpersonal/Peer Influence(IP)	3.8177	0.4279	-0.4314	2.4016
External Influence(EI)	3.8802	0.4337	-0.4168	2.5353
Perceived Usefulness(PU)	4.1367	0.4603	-0.5038	2.7686
Perceived Ease Of Use(PEOU)	3.9014	0.4393	-0.6339	3.1864
Compatibility(CP)	4.0305	0.4533	-0.8880	4.0649
Self-Efficacy(SE)	3.5797	0.4052	-0.2955	2.0911
Facilitating Conditions(FC)	3.3268	0.3811	-0.1721	2.0071
Perceived Behavioral Control(PBC)	3.3704	0.3800	-0.0898	2.2295
Attitude Towards Use(ATU)	4.3261	0.4843	-0.1404	2.7722
Subjective Norms(SN)	3.7773	0.4237	-0.1130	2.1069
Intention To Use(ITU)	4.2210	0.4721	-0.6699	3.2204

4.4 Model Validation

4.4.1 Reliability Analysis

Reliability of measurement items, convergent, and discriminant validity were analyzed using Confirmatory Factor Analysis(CFA) (Gefen & Straub, 2005). The standardized loadings for all measurement items, composite reliability, and average variance extracted of constructs are shown in Table 4.3.

Composite Reliability (CR) which was used to assess reliability measures the degree to which a model's dimensions are free from arbitrary error and therefore produce reliable outcomes (Mohd Suki & Ramayah, 2010). The CR of the constructs in the model ranged from 0.5291 to 0.8870 with the suggested cut-off point being 0.70 (Nunnally, 1994). External Influence (CR=0.6460), Facilitating Conditions (CR=0.5291), and Attitude towards Use (CR=0.5291) had a reliability of below 0.70 which may be due to the sample size.

Table 4.3: Reliability and Factor Loadings

Constructs /Measurement Items	**Standardized Loadings**	**CR**	**AVE**
Interpersonal/Peer Influence(IP)		**0.7659**	**0.5348**
interpersonal1	0.8989		
interpersonal2	0.4964		
interpersonal3	0.7416		

External Influence(EI)		**0.6460**	**0.3821**
externalinfluence1	0.5025		
externalinfluence2	0.6659		
externalinfluence3	0.6709		
Perceived Usefulness(PU)		**0.8148**	**0.5258**
puse1	0.6578		
puse2	0.8127		
puse3	0.7525		
puse4	0.6661		
Perceived Ease Of Use(PEOU)		**0.7579**	**0.4448**
peou1	0.5547		
peou2	0.5663		
peou3	0.7900		
peou4	0.7258		
Compatibility(CP)		**0.8632**	**0.6780**
compatibility1	0.7915		
compatibility2	0.8450		
compatibility3	0.8327		
Self-Efficacy(SE)		**0.8870**	**0.7244**
selfefficacy1	0.7811		
selfefficacy2	0.9263		
selfefficacy3	0.8398		
Facilitating Conditions(FC)		**0.5291**	**0.4449**
facilitating1	0.8075		
facilitating2	0.7929		
facilitating3	-0.2324		
Perceived Behavioral Control(PBC)		**0.7839**	**0.5574**
pbc1	0.6120		
pbc2	0.9428		
pbc3	0.6395		
Attitude Towards Use(ATU)		**0.6251**	**0.5633**
attitude1	0.8053		
attitude2	-0.6417		
attitude3	0.7384		
attitude4	0.8047		
Subjective Norms(SN)		**0.8082**	**0.5855**
sn1	0.7435		

sn2	0.8419		
sn3	0.7036		
Intention To Use(ITU)		**0.7621**	**0.5212**
intention1	0.6963		
intention2	0.8455		
intention3	0.6032		

4.4.2 Validity Analysis

Factor loadings and average variance extracted (AVE) were used to evaluate convergent validity. AVE was calculated by squaring all the factor loadings of each measurement item in a construct, adding them, and then dividing by the number of measurement items of the construct. Fornell & Larcker (1981) elucidates that factor loading and AVE of 0.5 are required for convergent validity. All factor loadings apart from facilitating3 (-0.2324) and attitude2 (-0.6417) were above the recommendation. The loading attitude2 had been framed on the negative side "*Using Big Data to generate statistics would be a foolish idea*" hence the sign was reversed.

Discriminant validity is measured to check the degree to which constructs in a model are different (Teo & Schalk, 2009). It was evaluated using AVE and comparing the correlation of the constructs with the square root of AVEs (diagonal elements in bold) as shown in table 4.4. For data to fulfill discriminant validity, the diagonal elements in bold should be less than the off-diagonal elements(Pearson's correlation of the constructs) in the corresponding rows and columns and AVE should be greater than 0.5 (Fornell & Larcker, 1981). External Influence (0.3821), Perceived Ease of Use (0.4448), and Facilitating Conditions (0.4449) had an AVE of less than 0.5. For convergence in the final model EI, PEOU, and FC were dropped since they have an AVE of less than 0.5.

Table 4.4: Constructs Correlation Matrix

	IP	EI	PU	PEOU	CP	SE	FC	PBC	ATU	SN	ITU
IP	**0.7313**										
EI	0.6132	**0.61814**									
PU	0.5412	0.5665	**0.7251**								
PEOU	0.2312	0.3797	0.4666	**0.6669**							
CP	0.5148	0.6052	0.6950	0.5352	**0.8234**						
SE	0.2739	0.3399	0.3865	0.5715	0.4579	**0.8511**					
FC	0.4157	0.4405	0.1516	0.2457	0.1642	0.4589	**0.6670**				
PBC	0.3662	0.4265	0.3553	0.6107	0.4152	0.8440	0.6263	**0.7466**			
ATU	0.5085	0.5582	0.6098	0.2201	0.8232	0.2514	0.0774	0.2233	**0.7505**		
SN	0.6871	0.5691	0.5315	0.3427	0.5611	0.5135	0.5647	0.6069	0.5915	**0.7652**	
ITU	0.6739	0.5213	0.6992	0.2685	0.7786	0.4230	0.3216	0.4467	0.7396	0.7271	**0.7219**

4.4.3 Structural Equation Model(SEM) Evaluation

This is a procedure to evaluate models and it consists of various methods such as paths, confirmatory factor, structural relation, and covariance structure analysis (Hair et al., 2006). For this study, SEM was used to evaluate the path analysis of the TAM-based model.

Best Fitting Model

For the model to be passed as fit to be used for SEM, it has to fulfill the cut-off points of various indices of the best fitting model. For this study, the Comparative Fit Index (CFI) which gauges the global fit Gerbing & Anderson (1993), and Normed Fit Index (NFI) which gauges the share by which a model's fit is enhanced compared to the base model Hair et al. (2006) were used to evaluate the model fit.

Table 4.5: Best-Fitting Model Results

Fit Indices	Value	Recommended Benchmark	References
CMIN (X^2)	1229.217 P=0.00	Non-significant	(Kline, 2010)
DF	583		
CMIN (X^2)/DF	2.108	<3	(Kline, 2010)
CFI (Comparative Fit Index)	0.901	≥0.9	(Bentler & Bonett, 1980)
NFI (Normed Fit Index)	0.903	≥0.9	(Bentler & Bonett, 1980)

The CMIN (X^2) is significant (p=0.000) and hence failed to satisfy the recommended benchmark of an acceptable fit. Hair et al. (2006) notes that CMIN (X^2) is affected by sample size differences and since the sample used in the study was 85 respondents may explain the failure to satisfy the recommended fit. The CFI and NFI values passed the minimum values of acceptable fit.

Analysis of Paths

The SEM model was drawn in STATA and the coefficients of paths between different dimensions were estimated. The path coefficients **(β)** represent standardized regression coefficients.

Table 4.6: Results of Hypotheses Test

Hypothesis	Path	β	S.E.	z	P-value	Result of the Hypothesis
H1	SN <--- IP	0.730	0.098	7.430	0.000	Supported
H2	SN <--- EI					Dropped from the model
H3	ATU <--- PU	0.579	0.163	3.560	0.000	Supported
H4	ATU <--- PEOU					Dropped from the model
H5	ATU <--- CP	0.529	0.161	3.280	0.001	Supported
H6	PBC <--- SE	0.951	0.056	17.090	0.000	Supported
H7	PBC <--- FC					Dropped from the model
H8	ITU <--- SN	0.424	0.215	1.970	0.049	Supported
H9	ITU <--- ATU	0.753	0.132	5.700	0.000	Supported
H10	ITU <--- PBC	0.116	0.138	0.841	0.400	Not Supported

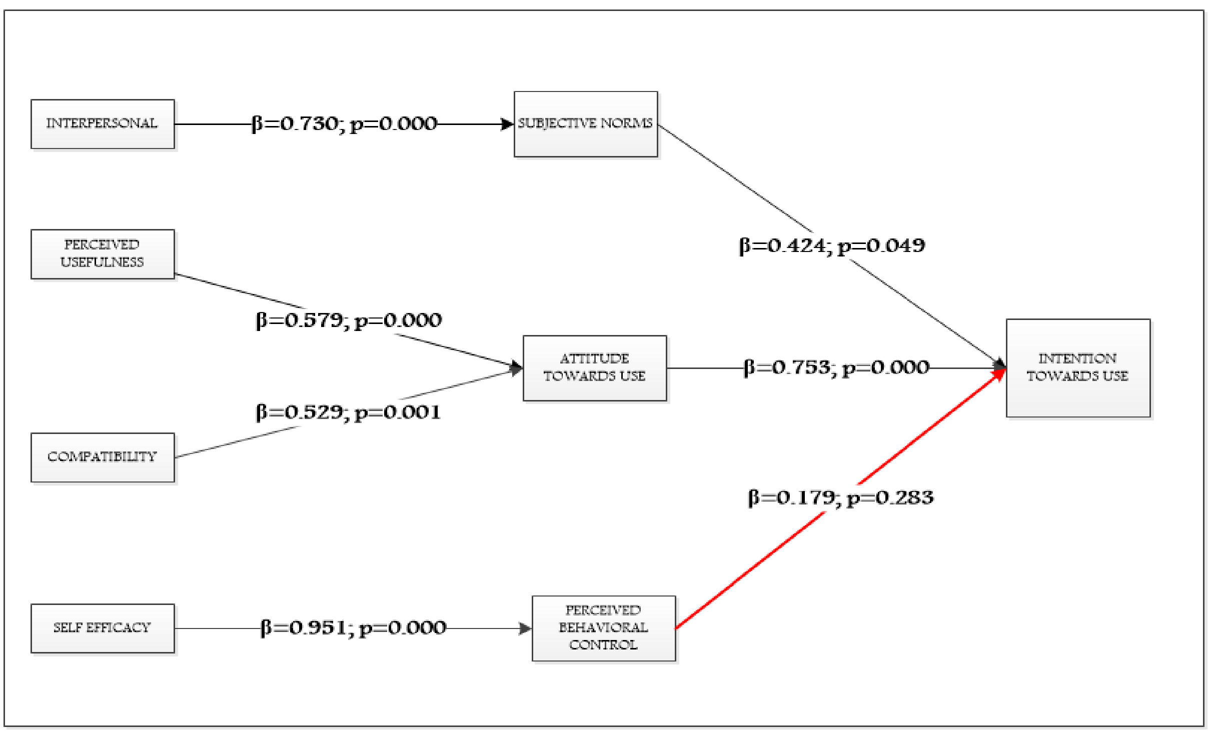

Figure 4.15: The Structural Model

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

This chapter details the achievements, conclusions, recommendations, and areas of further research based on the study conducted.

5.1 Achievements

Objective 1: Determine how Big Data is being used in research institutes.

This objective was evaluated both quantitatively from the survey conducted and qualitatively from the review of previous literature. Qualitatively, it was noted that the use of Big Data in Statistics is being advocated from the global level. Consequently, National Statistical institutes are using exhaust data, digital content, and sensing data from web scraping, Google maps, call detail records, satellite, and Twitter. Some of the projects noted include Infection Prevention and Control Online, Developing Water Accounts, Tourism Monitoring, Subjective Wellbeing, Movements across Borders, and Complementing the National Agriculture Census.

From the survey, it was noted that almost half of the institutes have a Big Data Strategy in place and foresee a role of Big Data in their work. The research institutes have existing potential Big Data sources that they can use in their work. At the individual's level, staff in data-related fields are of opinion that Big Data can solve data problems in their institutes and agree with the idea that Big Data can enhance and complement official statistics.

Objective 2: Establish the risks and challenges of using Big Data statistics.

The challenges of using Big Data in statistics were evaluated both quantitatively from the survey conducted and qualitatively. Previous literature noted that legal and regulatory issues; gaining access to data; gaining access to associated methodology and metadata; establishing dataset quality; establishing suitability for purpose; Institutional change management; and ensuring inter-organizational collaboration and common standards as challenges (Kitchin, 2015).

The survey conducted confirmed (Kitchin, 2015) findings that legal and regulatory issues; gaining access to data; gaining access to associated methodology and metadata; establishing dataset quality are the main challenges of using Big Data in statistics.

The risks of using Big Data in statistics were evaluated both quantitatively from the survey conducted and qualitatively. Previous literature highlighted inconsistent access and continuity; privacy breaches and data security; resource constraints and cut-backs; the resistance of Big Data providers and populace; privatization and competition, damage to reputation and losing public trust; and mission drift as the risks of using Big Data in Statistics(Kitchin, 2015).

The survey conducted confirmed (Kitchin, 2015) findings that inconsistent access and continuity; privacy breaches and data security; resource constraints and cut-backs; and resistance of Big Data providers and the populace as the most prominent risks.

Objective 3: Identify the determinants of adoption of Big Data in Statistics among research institutes.

The results demonstrated that the adoption of Big Data in statistics can be explained in terms of interpersonal influence, external influence, perceived usefulness, perceived ease of use, compatibility, self-efficacy, facilitating condition, and subjective norm. The study concludes that interpersonal/peer influence, perceived usefulness, compatibility, self-efficacy, perceived behavioral control, subjective norms, attitude towards the use of Big Data, and intention to use big data are the most significant determinants of the use of Big Data in statistics.

Objective 4: Validate the research model using Structural Equation Modeling (SEM).

The TAM-based model was validated with Perceived Usefulness (PU) and Compatibility (CP) validated to have a positive influence on Attitude towards Use (ATU) of a technology. Interpersonal/Peer Influence (IP) is validated to have a positive influence on Subjective Norms (SN). Self-Efficacy (SE) is validated to have a positive influence on Perceived Behavioral Control (PBC). Attitude towards Use (ATU) of Big Data in statistics and Subjective Norms (SN) are validated to have a positive influence on Intention to Use (ITU) Big Data in statistics. Perceived Behavioral Control (PBC) has an insignificant positive influence on Intention to Use (ITU) Big Data in statistics. The following is thus the validated model.

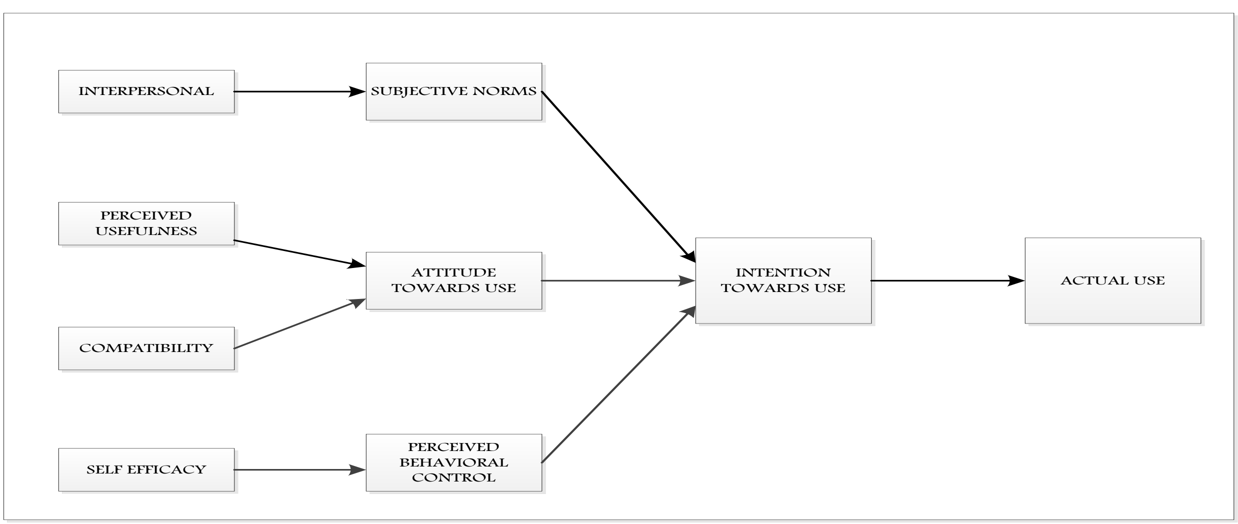

Figure 5.1: Validated Model

5.2 Conclusion

With Kenya and the rest of developing countries lagging with the number of statistical indicators being tracked, using Big Data is a necessity for evidence-based decision making since it is generated at a lower cost, greater frequency, and with a wider spatial distribution. Research institutions that are hubs for innovations and inventions must embrace Big Data to complement their functions.

This paper established that data professionals in research institutes agree that Big Data can complement traditional sources of data to generate statistics and are ready to adopt it. This calls for the training of employees to acquire the requisite skills and enhance self-efficacy which is a key determinant of adoption.

The challenges of legal and regulatory issues, gaining access to data and associated methodology and metadata, and establishing Big Data quality were noted to be the main challenges hindering the adoption and should be resolved. This calls for policy-makers to create an enabling legal and regulatory environment. Data sharing policies and agreements should be developed to handle the risk of inconsistent access of Big Data. Enforcement of non-disclosure agreements and a clear definition of personal data should be done to deal with privacy breaches.

The study established that external influence and subjective norms, perceived usefulness, compatibility, attitude towards use, and self-efficacy are the key factors influencing the adoption of Big Data in Statistics.

5.3 Recommendations

This study recommends research institutes undertake a radical shift in statistical methodology for Big Data to gain ground in statistics (Scannapieco et al., 2013). The current statistical methodologies of sampling and analysis should be improved to handle Big Data. Sensitization and capacity building of all stakeholders regarding the opportunities of Big Data should be done for it will improve the attitude towards use which is a key determinant of the use of Big Data in statistics. Legal issues which are facilitating conditions should be resolved by full implementation of the data protection act of 2019 and allocation of enough resources to Big Data projects. Training on new methodology and tools to handle Big Data should be enhanced to enhance self-efficacy. Advocacy should be enhanced since external influence and subjective norms play a significant role in the adoption.

5.4 Research Assessment

This research is assessed using the seven key questions formulated by (Whetten, 1989).

What is new? Does the thesis make a significant, value-added contribution to the current thinking?

The information in this study will guide research institutes on the factors that they should mainly focus on to promote high adoption of Big Data in statistics. This study validates the path analysis of the core dimensions of TAM i.e. Perceived Usefulness (PU), Attitude towards Use (ATU), and Intention to Use (ITU). It also establishes Compatibility (CP) and Self-Efficacy (SE) as key dimensions in the adoption of Big Data in statistics. This study is consistent with other studies on the adoption of other technologies using TAM.

So what? How will the research change the current thinking and practice?

The study will guide research institutes on the key determinants to prioritize when adopting Big Data in statistics. This study also highlights the challenges and risks of the use of Big Data in statistics which will guide research institutes as they adopt Big Data in their operations. Policy-makers will be guided on opportunities of Big Data to track real-time indicators. It also seeks to stir up debate in the Statistics field which will undergo major changes due to Big Data technology.

Are the underlying logic and supportive evidence compelling?

This study is based on a TAM-based model which is widely used and extended to assess technology adoption. The results of this study agree with past studies. The findings also show that Technology Acceptance Model is robust and can be extended and applied to any technology adoption. The challenges and risks are sourced from respected authors and validated by the respondents of this study which demonstrates consistency.

Why now? Is it of interest to the people?

The adoption of Big Data in statistics is being advocated and the findings will guide research institutes in their adoption. The opportunities presented by Big Data need to be explored to fill the data gaps and enable Kenya to track more sustainable development goals (SDGs).

Who else including academic researchers is interested in this study?

The first beneficiary of this study is research institutes. Other interested parties are statisticians, statistical institutes, and National Statistical Offices and Policy-makers.

5.5 Further Research

A future study should focus on a statistics domain area like tourism statistics or consumer price index and seek to generate statistics using Big Data sources and compare with the current surveys and administrative data sources. A study based on more than the eleven variables used here should be carried out. Another model can also be adopted to find whether it would be more suitable. Further studies should increase the sample size and assess all the institutes that constitute the National Statistical System.

REFERENCES

About PARIS21. (2021). Paris 21. https://paris21.org/about-paris21

About—UN-CEBD. (2021). https://unstats.un.org/bigdata/about/index.cshtml

Abrahams, D. (2010). Technology Adoption in Higher Education: A Framework for Identifying and Prioritising Issues and Barriers to Adoption of Instructional Technology. *Journal of Applied Research in Higher Education, 2*, 34–49. https://doi.org/10.1108/17581184201000012

African Union. (2021). *Pan-African Institute for Statistics (STATAFRIC) | African Union*. https://au.int/ea/statistics/statafric

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (1980). Understanding Attitudes and Predicting Social Behavior Prentice-Hall. In *Englewood Cliffs, NJ, Prentice-Hall* (Vol. 278).

Al-Raqadi, A., Abdul Rahim, A., Masrom, M., & Al Riyami, B. (2015). An Assistive and Research Framework Methodology for Ships' Upkeep and Repair Organisational Learning Performance. *Asian Social Science*, *11*. https://doi.org/10.5539/ass.v11n18p246

Badiee, S., Jütting, J., Appel, D., Klein, T., & Swanson, E. (2017). *The role of national statistical systems in the data revolution*. OECD.

Baker, J. (2011). The Technology–Organization–Environment Framework. In *Information Systems Theory* (pp. 231–245). https://doi.org/10.1007/978-1-4419-6108-2_12

Bandura, A. (1986). *Fearful expectations and avoidant actions as coeffects of perceived self-inefficacy*.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588.

Beselga, D., & Alturas, B. (2019). *Using the Technology Acceptance Model (TAM) in SAP Fiori* (pp. 575–584). https://doi.org/10.1007/978-3-030-16181-1_54

Big data. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Big_data&oldid=976755644

Consoli, D. (2012). Literature analysis on determinant factors and the impact of ICT in SMEs. *Procedia-Social and Behavioral Sciences, 62*, 93–97.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.*, *13*, 319–340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science, 35*(8), 982–1003. JSTOR.

Dearing, J., & Cox, J. (2018). Diffusion Of Innovations Theory, Principles, And Practice. *Health Affairs, 37*, 183–190. https://doi.org/10.1377/hlthaff.2017.1104

Demunter, C. (2017). Tourism statistics: Early adopters of big data. *Luxembourg: EUROSTAT*.

Dishaw, M. T., & Strong, D. M. (1999). Extending the technology acceptance model with the task–technology fit constructs. *Information & Management*, *36*(1), 9–21. https://doi.org/10.1016/S0378-7206(98)00101-3

Economic, U., & Council, S. (2016). *Report of the Global Working Group on Big Data for Official Statistics*.

Events—UN-CEBD. (2021). https://unstats.un.org/bigdata/events/index.cshtml

Florescu, D., Karlberg, M., Reis, F., Castillo, P. R. D., Skaliotis, M., & Wirthmann, A. (2014). *Will 'big data' transform official statistics?* 12.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50. JSTOR. https://doi.org/10.2307/3151312

Fritz, S., See, L., Carlson, T., Haklay, M., Oliver, J., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., & West, S. (2019). *Citizen science and the United Nations Sustainable Development Goals*. *2*, 922–930. https://doi.org/10.1038/s41893-019-0390-3

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*(1), 5.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. *Sage Focus Editions*, *154*, 40–40.

Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*, *19*(2), 213–236. JSTOR. https://doi.org/10.2307/249689

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. (2006). *Multivariate data analysis. Uppersaddle River*. NJ: Pearson Prentice Hall.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, *65*(23), 2276–2284.

Kitchin, R. (2015). The opportunities, challenges, and risks of big data for official statistics. *Statistical Journal of the IAOS*, *31*, 471–481. https://doi.org/10.3233/SJI-150906

Kline, R. (2010). Principles And Practice Of Structural Equation Modeling. In *Guilford Press*.

KNBS. (2019). Kenya Population and Housing Census Volume IV: Distribution of Population by Socio-Economic Characteristics. *Nairobi, Kenya: Kenya National Bureau of Statistics, IV*.

KNBS. (2020). *Kenya Strategy for Development of Statistics 2019/20—2022/23*. http://www.knbs.or.ke/

KNBS Mandate. (2020). *Kenya National Bureau of Statistics*. https://www.knbs.or.ke/?page_id=85

Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.

Landefeld, S. (2014). Uses of big data for official statistics: Privacy, incentives, statistical challenges, and other issues. *International Conference on Big Data for Official Statistics Beijing, China*, 28–30.

Letouzé, E., & Jütting, J. (2015). Official statistics, big data, and human development. *Data-Pop Alliance White Paper Series. Data-Pop Alliance, New York*.

MacFeely, S. (2018). *Big Data and Official Statistics* (pp. 25–54). https://doi.org/10.4018/978-1-5225-7077-6.ch002

MacFeely, S. (2019). The big (data) bang: Opportunities and challenges for compiling SDG indicators. *Global Policy*, *10*, 121–133.

Mohd Din, S. H., akub, M., & Chellamuthu, D. (2017). *Big Data Innovate Official Statistics For Better Decision Making*.

Mohd Suki, N., & Ramayah, T. (2010). User Acceptance of the E-Government Services in Malaysia: Structural Equation Modelling Approach. *Interdisciplinary Journal of Information, Knowledge, and Management*, *5*. https://doi.org/10.28945/1308

Mugenda, O. M., & Mugenda, A. G. (2003). *Research Methods*. Acts Press, Nairobi.

NSS. (2017). Paris 21. https://paris21.org/national-statistical-system-nss

Nunnally, J. C. (1994). *Psychometric theory 3E*. Tata McGraw-hill education.

Oates, B. (2005). *Researching Information Systems and Computing*.

OECD. (2002). *Measuring the Non-Observed Economy*. OECD Publications Service. http://www.oecd.org/dataoecd/9/20/1963116.pdf

OECD. (2013). *Strengthening National Statistical Systems to Monitor Global Goals, OECD and Post-2015 Reflections, Element 5, Paper 1*. https://www.oecd.org/dac/POST-2015%20P21.pdf

Official statistics. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Official_statistics&oldid=962182066

OLeary, D., & Storey, V. C. (2020). Discovering and Transforming Exhaust Data to Realize Managerial Value. *Communications of the Association for Information Systems*, *47*(1), 11.

Oliveira, T., & Martins, M. R. (2011). Literature Review of Information Technology Adoption Models at Firm Level. *1566-6379*, *14*.

PARIS21. (2015). *A road map for a country-led data revolution.* https://doi.org/10.1787/9789264235250-fr

Rahman, N. (2016). *Factors affecting big data technology adoption*.

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, *19*(4), 1–34.

Sangokoya, D., Letouzé, E., Barrett, L., & Manske, J. (2015, November 12). *"Opportunities and Requirements for Leveraging Big Data for Official Statistics and the Sustainable Development Goals in Latin America"*. Data-Pop Alliance.

https://datapopalliance.org/item/opportunities-and-requirements-for-leveraging-big-data-for-official-statistics-and-the-sustainable-development-goals-in-latin-america/

Saunders, M., Lewis, P., & Thornbill, A. (2000). *Research Methods for Business Studies*.

Scannapieco, M., Virgillito, A., & Zardetto, D. (2013). Placing big data in official statistics: A big challenge. *New Techniques and Technologies in Statistics*.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Struijs, P., Braaksma, B., & Daas, P. J. H. (2014). *Official statistics and Big Data*.

Tam, S.-M., & Clarke, F. (2015). Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics. *International Statistical Review*, *83*. https://doi.org/10.1111/insr.12105

Tam, S.-M., & Halderen, G. (2020). The five V's, seven virtues, and ten rules of big data engagement for official statistics. *Statistical Journal of the IAOS*, *36*, 1–11. https://doi.org/10.3233/SJI-190595

Teo, T., & Schalk, P. (2009). Understanding Technology Acceptance in Pre-Service Teachers: A Structural-Equation Modeling Approach. *Asia-Pacific Education Researcher - ASIA-PAC EDUC RES*, *18*. https://doi.org/10.3860/taper.v18i1.1035

Tornatzky, L. G., & Klein, K. J. (1982). Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. *IEEE Transactions on Engineering Management*, *EM-29*(1), 28–45. https://doi.org/10.1109/TEM.1982.6447463

Turner, R. (2007). Diffusion of Innovations, 5th edition, Everett M. Rogers. Free Press, New York, NY (2003), 551 pages. *Journal of Minimally Invasive Gynecology*, *14*, 776. https://doi.org/10.1016/j.jmig.2007.07.001

UN. (2013). *The Millennium Development Goals Report 2013* (p. 68). UN. https://www.un.org/millenniumgoals/pdf/report-2013/mdg-report-2013-english.pdf

UN GWG on Big Data for Official Statistics. (2017). *Bogota Declaration*. https://unstats.un.org/unsd/bigdata/conferences/2017/Bogota%20declaration%20-%20Final%20version.pdf

UN Statistical Commission. (2017). Cape Town global action plan for sustainable development data. *Adopted by the UN Statistical Commission at Its 48th Session*. https://unstats.un.org/sdgs/hlg/Cape_Town_Global_Action_Plan_for_Sustainable_Development_Data.pdf

UNECE. (2014). *Fundamental Principles of Official Statistics | UNECE*. https://unece.org/statistics/fundamental-principles-official-statistics

UNSD — United Nations Statistical Commission. (2021). https://unstats.un.org/unsd/statcom

Venkatesh, V., Thong, J., & Xu, X. (2016). Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. *Journal of the Association for Information Systems*, *17*, 328–376. https://doi.org/10.17705/1jais.00428

Weng, F., Yang, R.-J., Ho, H.-J., & Su, H.-M. (2018). A TAM-Based Study of the Attitude towards Use Intention of Multimedia among School Teachers. *Applied System Innovation*, *1*, 36. https://doi.org/10.3390/asi1030036

Whetten, D. (1989). What Constitutes A Theoretical Contribution? *Academy of Management Review*, *14*, 490–495. https://doi.org/10.2307/258554

White Baker, E., Al-Gahtani, S., & Hubona, G. (2007). The effects of gender and age on new technology implementation in a developing country: Testing the theory of planned behavior (TPB). *IT & People*, *20*, 352–375. https://doi.org/10.1108/09593840710839798

APPENDICES

APPENDIX I: INTRODUCTORY LETTER

Dear Respondent,

RE: DATA COLLECTION FOR RESEARCH

My name is Allan Gathuru Wairimu, a student pursuing a Master's degree in Information technology management at the University of Nairobi. I am collecting data for my project which seeks to assess the adoption of Big Data in statistics.

I humbly request you to contribute to my research by filling the attached questionnaire following the instructions. Your feedback is of high value to my study will add a lot of value to my study.

All the data provided will be strictly be used for this research and will be handled strictly confidentially. All your details will be anonymized during the report generation.

Thank you in advance,
Yours Sincerely,

Allan Gathuru Wairimu
P54/35452/2019

APPENDIX II: QUESTIONNAIRE

SECTION 1: DEMOGRAPHIC INFORMATION

1. What is your profession

   Statistician [ ] Data Scientist [ ] Data Analyst [ ] Data Manager [ ]

2. Kindly indicate your gender

   Male [ ] Female [ ]

3. What is your age bracket?

   18 - 24 [ ] 25-34 [ ] 35-44 [ ] 45-60 [ ] Above 60 [ ]

4. How many years of experience do you have?

   0 - 5 [ ] 6-10 [ ] 11-20 [ ] Above 20 [ ]

5. Does your organization has any Big Data Strategy?

   Yes [ ] No [ ]

6. Does your organization foresee any role for Big Data in the execution of your work?

   Yes [ ] No [ ]

7. Are there existing or potential Big Data Sources for your organization?

   Yes [ ] No [ ]

8. In Your Opinion, does Big Data has the potential to alleviate data challenges in your organization?

   Yes [ ] No [ ]

9. Do you agree with the notion that Big Data can be used to supplement official statistics?

   Yes [ ] No [ ]

10. What do you consider to be the major risks to the use of Big Data in Official statistics?
    a)……………………………………………………………………………………
    b)……………………………………………………………………………………
    c)……………………………………………………………………………………

d)……………………………………………………………………………………

11. What are the major obstacles to the use of Big Data in official statistics in your organization?

a)……………………………………………………………………………………

b)……………………………………………………………………………………

c)……………………………………………………………………………………

SECTION 2: INTERPERSONAL/PEER INFLUENCE	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
My peers/colleagues/friends think that I should use Big Data when generating statistics.					
People I know think that using Big Data when generating statistics is a good idea.					
People I know influence me to try using Big Data when generating statistics					

SECTION 3: EXTERNAL INFLUENCE	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I read/ saw news reports that using Big Data analytics was a good way to generate statistics.					
The popular press depicted a positive sentiment for using Big Data when generating statistics.					
Mass media reports influenced me to try using Big Data when generating statistics.					

SECTION 4: PERCEIVED USEFULNESS	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Using Big Data when generating statistics would improve my performance.					
Using Big Data when generating statistics would improve my productivity.					
Using Big Data when generating statistics would enhance my effectiveness.					
I would find Big Data when generating statistics useful.					

SECTION 5: PERCEIVED EASE OF USE	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
Learning to use Big Data when generating statistics would be easy for me.					
I would find it easy to get more indicators using Big Data when generating statistics.					
It would be easy for me to become skillful at generating statistics using Big Data.					
I would find Big Data analytics tools for generating statistics easy to use.					

SECTION 6: COMPATIBILITY	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
Using Big Data when generating statistics will fit well with the way I work.					
Using Big Data when generating statistics will fit into my work style.					
The setup of Big Data for generating statistics will be compatible with the way I work.					

SECTION 7: SELF EFFICACY	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I would feel comfortable using Big Data when generating statistics on my own.					
I would be able to use Big Data when generating statistics reasonably well on my own.					
I would be able to use Big Data when generating statistics even if there was no one around to help me.					

SECTION 8: FACILITATING CONDITIONS	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The resources required to generate statistics from Big Data are available to me.					
I have access to hardware, software, and services needed to generate statistics from Big Data.					
I am constrained by the lack of resources needed to generate statistics from Big Data.					

SECTION 9: PERCEIVED BEHAVIOURAL CONTROL	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I would be able to generate statistics from Big Data well.					
Using Big Data analytics to generate statistics is entirely within my control.					
I have the resources, knowledge, and ability to generate statistics from Big Data.					

SECTION 10: ATTITUDE TOWARDS USE	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Using Big Data to generate statistics would be a good idea					
Using Big Data to generate statistics would be a foolish idea.					
I like the idea of using Big Data to generate statistics.					
Using Big Data to generate statistics would be a pleasant experience.					

SECTION 10: SUBJECTIVE NORMS	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
People (peers and experts) important to me support my use of Big Data to generate statistics.					
People who influence my behavior want me to use Big Data to generate statistics in addition to other data sources.					
People whose opinions I value prefer that I use Big Data to generate statistics.					

SECTION 11: INTENTION TO USE	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I intend to use Big Data to generate statistics.					
I will likely use Big Data to generate statistics.					
I expect to use Big Data to generate statistics in the future.					

APPENDIX III: CONSTRUCTS AND CORRESPONDING ITEMS

CONSTRUCTS	Items	Item Descriptions
INTERPERSONAL/ PEER INFLUENCE	interpersonal1	My peers/colleagues/friends think that I should use Big Data when generating statistics.
	interpersonal2	People I know think that using Big Data when generating statistics is a good idea.
	interpersonal3	People I know influence me to try using Big Data when generating statistics
EXTERNAL INFLUENCE	externalinfluence1	I read/ saw news reports that using Big Data analytics was a good way to generate statistics.
	externalinfluence2	The popular press depicted a positive sentiment for using Big Data when generating statistics.
	externalinfluence3	Mass media reports influenced me to try using Big Data when generating statistics.
PERCEIVED USEFULNESS	pu1	Using Big Data when generating statistics would improve my performance.
	pu2	Using Big Data when generating statistics would improve my productivity.
	pu3	Using Big Data when generating statistics would enhance my effectiveness.
	pu4	I would find Big Data when generating statistics useful.
PERCEIVED EASE OF USE	peou1	Learning to use Big Data when generating statistics would be easy for me.
	peou2	I would find it easy to get more indicators using Big Data when generating statistics.
	peou3	It would be easy for me to become skillful at generating statistics using Big Data.
	peou4	I would find Big Data analytics tools for generating statistics easy to use.
COMPATIBILITY	compatibility1	Using Big Data when generating statistics will fit well with the way I work.
	compatibility2	Using Big Data when generating statistics will fit into my work style.
	compatibility3	The setup of Big Data for generating statistics will be compatible with the way I work.
SELF EFFICACY	selfefficacy1	I would feel comfortable using Big Data when generating statistics on my own.
	selfefficacy2	I would be able to use Big Data when generating statistics reasonably well on my own.
	selfefficacy3	I would be able to use Big Data when generating statistics even if there was no one around to help me.
	facilitating1	The resources required to generate statistics from Big Data are available to me.

FACILITATING CONDITIONS	facilitating2	I have access to hardware, software, and services needed to generate statistics from Big Data.
	facilitating3	I am constrained by the lack of resources needed to generate statistics from Big Data.
PERCEIVED BEHAVIOURAL CONTROL	pbc1	I would be able to generate statistics from Big Data well.
	pbc2	Using Big Data analytics to generate statistics is entirely within my control.
	pbc3	I have the resources, knowledge, and ability to generate statistics from Big Data.
ATTITUDE TOWARDS USE	attitude1	Using Big Data to generate statistics would be a good idea
	attitude2	Using Big Data to generate statistics would be a foolish idea.
	attitude3	I like the idea of using Big Data to generate statistics.
	attitude4	Using Big Data to generate statistics would be a pleasant experience.
SUBJECTIVE NORMS	sn1	People (peers and experts) important to me support my use of Big Data to generate statistics.
	sn2	People who influence my behavior want me to use Big Data to generate statistics in addition to other data sources.
	sn3	People whose opinions I value prefer that I use Big Data to generate statistics.
INTENTION TO USE	intention1	I intend to use Big Data to generate statistics.
	intention2	I will likely use Big Data to generate statistics.
	intention3	I expect to use Big Data to generate statistics in the future.