RESEARCH PAPER

AFROSAI-E

Integrating Big Data in Public Sector Audit

December 2020

Contents

Acronyms

AFROSAI	African Organisation of Supreme Audit Institutions
AFROSAI-E	African Organisation of English-speaking Supreme Audit Institutions
CREFIAF	African Organisation of French-speaking Supreme Audit Institutions
HR	Human resources
ICBF	Institutional Capacity Building Framework
ICT	Information and communications technology
IFMIS	Integrated Financial Management Information System
INTOSAI	International Organisation of Supreme Audit Institutions
IS	Information system
ISSAI	International Standards of Supreme Audit Institutions
IT	Information technology
MDA	Ministries, departments and agencies
NGO	Non-governmental organisation
SAI	Supreme Audit Institution
SDG	Sustainable Development Goal
WGBD	(INTOSAI) Working Group on Big Data

EXECUTIVE SUMMARY

New analytics tools and methods are expanding the possible ways in which SAIs can derive value from existing data from audited entities and from freely available external information sources like non-governmental organisations (NGOs) or commercial data sources. While traditional data analytics has generally targeted structured data, like general ledgers that can be easily parsed and analysed, advances in analytics methods now allow examination of more varied data types. These advances could allow Supreme Audit Institutions (SAIs) to attempt a wider variety of audit subject matters, issue better reports and make better recommendations. With these advancements as a backdrop, the African Organisation of Supreme Audit Institutions (AFROSAI) commissioned a paper to explore how to integrate big data into public audit. The following are the key messages to emerge from this study:

Message 1: Big data or not so big data – it does not matter

'Big data' has become a buzzword and is often being used to describe any form of data analysis. Whether we are dealing with big data analyses or simply analysing large amounts of data depends on the definition. Simply put, big data is large in *volume* (terabytes, even petabytes), complex in *format* (structured and unstructured formats like photos, maps, text, video, satellite data, sound etc.) and it is aggregated at a high speed (*velocity*), making regular processing power and software inadequate. More attributes have been added to the definition in recent years, like *value* – data has become a commodity in itself, either by providing a competitive advantage or as an asset, and *veracity* – that questions the reliability of the data. Most SAIs are analysing data to some extent in their audits – does it matter whether it is big data or not so big data? We think not. "Value" should be the key word for SAIs. If the data analytics that we are conducting adds value to our audits, our reports and the citizens, we want to encourage more of it.

Message 2: Governments are lagging behind with regard to data and digitalisation

Despite the focus on digitalisation at country level, African nations are still far behind developed or higher income countries in terms of digitalisation and automation. In many African countries:

- Key processes and reporting in the public sector are not yet fully digitalised.
- Government data is not readily available both within government and to the public.
- The data is not available in machine-readable format and is not reusable, not free of charge and not updated regularly.
- Data is not used actively in decision-making and for service delivery.
- The public sector is not governing, managing and valuing data in the same way as the other assets that are relevant for their success.

Supreme Audit Institutions (SAIs) could play an important role in informing and guiding governments regarding the benefits of structural reforms and openness of public data. SAIs could promote the

principle of availability and openness of data produced by the government if confidentiality laws or privacy concerns do not rule it out. This could contribute to greater transparency and support government accountability. SAIs could also promote openness of automated decision-making systems used by government according to principles of open source code and open data.

Message 3: We have to work smarter

The study reveals that within the SAIs, there is insufficient knowledge about the usage of data and data analytics in audit. It is customary for the information systems (IS) auditors to be charged with the task of capturing the data and analysing it. The study also suggests that many SAIs use the terms "IS audit" and "data analysis" interchangeably. Data analytics should not be limited to IS audit alone, as it can also enhance audit quality in financial and performance audits. Data analytics enables auditors to improve the risk assessment process, substantive procedures and tests of controls. It can involve very simple routines as well as complex models that produce high-quality projections. Auditors using such models need to understand them and exercise significant judgement in determining when and how they should be used.

Message 4: SAIs have to put data on the agenda

Despite the respondents' optimistic view of the SAIs' readiness for big data analytics, the SAIs in the region do not have the people, the processes, nor the technology to build capacities in big data analysis. Very few SAIs have assessed their internal capacities, which implies that there is not proper buy-in from the top. The decision to become a more data-driven SAI is a strategic one, hence it should be outlined in the SAI's strategy. Building capabilities and infrastructure to match is a project that needs buy-in from the entire organisation. We need to assess how to derive value from data analytics in audit, and to assign roles and responsibilities throughout the organisation. We need to build partnerships in and beyond our public sectors to ensure access to relevant data. We need to empower the people that are already in our ranks and have a plan for how to recruit for the future.

The SAIs should have a clear strategy for data analysis in all three types of audit, which will ensure buy-in throughout the organisation. The strategy should include the following:

- ✓ ***Direction and impact:*** The utilisation of data in audit has to have a purpose. The SAI should assess how data analytics can be integrated in the three types of audit and what its desired impact should be. The assessment should also identify risks that have to be addressed in order for the SAI to achieve its goals.
- ✓ ***People:*** To become more data driven, the SAI has to assign clear roles and responsibilities. The HR department has to assess whether the needed skills can be found in-house or whether recruitment is needed. Training programs for staff may have to be designed and implemented, for both auditors and expert data analysts.
- ✓ ***Processes:*** The SAI should identify stakeholders and form partnerships with both governmental institutions and external institutions who maintain relevant data repositories to ensure continuous access to all relevant data. Where there is lack of data, or government

fails to update or keep data in machine-readable and reusable formats, the SAI needs a strategy to find alternative data sources. The SAI should also include any shortcomings in government's handling of data in its reports and recommendations. Processes for data capture and transfer from the auditee/data partner to the office and distribution to relevant audit teams will have to be designed and implemented. Quality of data and veracity are of the utmost importance. The SAI needs processes to ensure that the data it uses in audit has passed quality control and is verified. Identifying areas in which the SAI will use data analytics in audit, will facilitate the process of data quality in these specific areas.

- ✓ ***Technology:*** Although we need tools fit for the purpose of data analytics, audit quality does not lie in the tools themselves but rather in the quality of data and the analyses and judgements of the auditors. Some SAIs may need to invest in data analysis software; although it is important to note that you can conduct quality analyses using Excel. However, through its assessment of its capabilities, the SAI may find that its infrastructure with regard to network, storage, processing power and analysis tools is not adequate. Investments in technology may be a significant cost, and a plan for scaling towards the desired level should be included in the SAI's strategy.

1 INTRODUCTION TO THE THEME – INTEGRATING BIG DATA IN PUBLIC SECTOR AUDIT

1.1 Background – INCOSAI XXIII

During the 23rd INTOSAI congress (INCOSAI XXIII) in Moscow in 2019, two resolutions concerning data and data analytics were drafted, in which SAIs are encouraged to respond effectively to opportunities brought by technological advancement.

The first is a call for SAIs to promote the principle of availability and openness of data, source code and algorithms.

Key statements from INCOSAI:

- ✓ More and more data generated by public administrations is being made available in open formats. This creates a new information environment for everyone and can contribute to greater transparency as well as supporting government accountability.
- ✓ SAIs could play an important role in informing and guiding governments regarding the benefits of public data openness. SAIs could promote the principle of availability and openness of data produced by the government if confidentiality laws or privacy concerns do not rule it out.
- ✓ SAIs could promote openness of automated decision-making systems used by government under principles of open source code and open data.
- ✓ SAIs should facilitate a discussion about the publication of SAIs' results in an open data format, when appropriate.[1]

The second resolution states that SAIs could aim to make better use of data analytics in audits, including adaptation strategies, such as planning for such audits, developing experienced teams for data analytics, and introducing new techniques into the practice of public audit.

Key statements:

- ✓ Use of data analytics in SAIs is a necessary innovation that makes data a resource for promoting the efficiency, accountability, effectiveness and transparency of public administration.
- ✓ The unique position of SAIs within the public sector enables them to collect large amounts of data from audited entities. Employment of big data analysis techniques through the auditing process enables SAIs:
    - ➢ To synthesise data obtained from different departments, fields, government levels and regions, which enables the synthesis of obtained data to find solutions to transversal government problems.

[1] INTOSAI. 2019. *XXIII INCOSAI – Moscow Declaration.* Vienna: INTOSAI.

- To combine on-site, off-site and mixed data collection approaches to provide regular data updates and enable real-time monitoring of overarching issues and major risk areas.

✓ SAIs could benefit from conducting big data analytical research at the preliminary stage of audits. It will shorten fieldwork time and allow for regular monitoring of follow-up.

✓ SAIs could benefit from extending research studies and enhancing their in-house methodological work to apply sound and appropriate research methods. SAIs could also engage academic institutions to conduct cooperative research.

✓ By strengthening cooperation among SAIs and relevant international organisations, INTOSAI can summarise big data audit experience and knowledge, develop relevant guidelines and research reports, and encourage SAIs to build up their big data audit capabilities.[2]

1.2 Context of adoption of the theme by AFROSAI

The 55th meeting of the AFROSAI Governing Board, held in Dakar on 29-30 August 2019, adopted this technical theme for the next AFROSAI General Assembly[3]. This study is therefore undertaken in fulfilment of the role that AFROSAI-E was given to develop a technical paper on theme 2: "Integrating Big Data in Public Sector Audit".

[2] INTOSAI. 2019. *XXIII INCOSAI – Moscow Declaration.* Vienna: INTOSAI.

[3] AFROSAI. 2020. Preparation of AFROSAI 15th General Assembly Technical Theme Reports. Dakar: AFROSAI.

2 CONTEXT OF THEME STUDY

2.1 Definition of big data

The term “big data” is used to describe the exponential growth and availability of data created by people, applications and smart machines, as well as large complex data sets that are beyond the capabilities of traditional data-processing applications. The proliferation of structured and unstructured data, combined with technical advances in storage, processing power and analytic tools, has enabled big data to become a competitive advantage for organisations that use it to gain insights into business opportunities and drive business strategies.

Big data is generated in organisations (e.g. transaction data, customer data), industries, societies (e.g. economic data, CCTV, government databases), nature (e.g. weather, seismic data, locations and maps), social media, market research and numerous other sources.

Traditionally, most data stored within organisations has been structured and maintained in relational databases. This structured data is organised and allows for repeatable queries. It is often easier to control due to defined ownership and vendor-supported database solutions. The use of unstructured data, however, is growing and becoming more common in organisations. This type of data is not confined to traditional structures or constraints. It is typically more difficult to manage due to its evolving and unpredictable nature, and it is usually sourced from large, disparate and often external data sources.

Big data is used for prediction by using a complex method of (big data) analytics to infer information from data sets from a variety of different sources. The data can be collected from various sensors, such as − but not limited to − internet clicks, satellite devices (e.g. GPS data), wearable devices, swipe cards, payment devices, health information and weather sensors.

2.1.2 Characteristics of big data

The characteristics of big data have been defined[4] by the following 7 “V”s:

- ✓ **V**olume: The amount of data being created is vast compared to traditional data sources.
- ✓ **V**ariety: Data comes from all types of formats, including data generated within an organisation as well as data created from external sources.
- ✓ **V**elocity: Data is being generated extremely quickly and continuously.
- ✓ **V**eracity: Data must be able to be verified based on both accuracy and context.

[4] Johnson, L. 2019, June: 4 Vs of Big Data & 2019, December: Three More Vs of Big Data. Isaca.org https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2019/volume-12/4-vs-of-big-data https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2019/volume-25/three-more-vs-of-big-data

- ✓ **V**ariability: Big data is extremely variable and always changing.
- ✓ **V**isualisation: Analytic results from big data are often hard to interpret; therefore, translating vast amounts of data into readily presentable graphics and charts that are easy to understand is critical to end-user satisfaction and may highlight additional insights.
- ✓ **V**alue: Organisations, societies and consumers can all benefit from big data. Value is generated when new insights are translated into actions that create positive outcomes.

Additional "V"s and further characteristics of big data have been defined but the above highlight the main agreed characteristics.

Risks associated with big data include poor data quality, inadequate technology, insufficient security and immature data governance practices[5].

2.1.1 Structured and unstructured data – an example

The Global Polio Eradication Initiative (GPEI) teamed biological science with technological science to eradicate type 2 polio by changing the approach to vaccination delivery and surveillance[6]. The last place on earth with type 2 polio was Borno state in northern Nigeria. Vaccination teams would travel from village to village in remote, rural regions of the countryside administering the polio vaccination. The problem was, they had to rely on hand-drawn maps, sketched from memory, and often missed whole villages. Thousands of children would go unvaccinated and this highly contagious disease would rage on. GPEI could only identify a polio outbreak once children started showing up paralyzed but had no means to predict where the disease would appear next. GPEI then discovered that by analysing data from location technology such as satellites and GPS tracking (unstructured data) and using mapping software to create highly detailed maps, they could spur eradication efforts. Vaccination teams were now armed with a holistic picture showing where outbreaks were, which villages had been missed and the number of teams that had to be deployed to contain it. These maps, along with (structured) data from monitoring sewage would provide vital information about where the virus was, and where to immunise people before the virus infected them. On 25 August 2020, Africa was declared polio free by the independent body, the Africa Regional Certification Commission[7].

2.2 Uses or applications of big data in audit

Digitalisation has added to the amount of data produced worldwide. Big data is increasing rapidly with a study in 2015 projecting a 40%[8] growth. Smart data analysis has become a challenging task in today's environment where disparate data sets are generated across the globe with enormous

[5] Bekker, A. 2018, March: The 'scary' seven: big data challenges and ways to solve them. Scnsoft.com
[6] Dickson, I. 2019, November: https://360.here.com/polio-eradication-rotary
[7] World Health Organisation Africa. 2020: https://www.afro.who.int/news/africa-eradicates-wild-poliovirus
[8] Al Nuaimi, E., Al Neyadi, H., Mohamed, N. & Al-Jaroodi, J. 2015. Applications of big data to smart cities. *Internet Serv Appl.*

volumes[9]. Many applications and uses of big data have been identified in both public and private sectors. Some examples of the use of big data analytics are the following:

- ✓ Analysis of data sets can find new correlations e.g. to spot business trends, prevent diseases, and combat crime.
- ✓ Predictive analytics, user behaviour analytics.
- ✓ To support in the implementation and provision of smart city services[10].
- ✓ Detection of fraud and medical malpractice[11].
- ✓ Risk identification in compliance and other audits[12].
- ✓ Use of big data in innovative applications for Sustainable Development Goals (SDGs) like poverty alleviation, health, improving the livelihood of people, agriculture, amongst others[13].
- ✓ Auditing the achievement of SDGs, e.g. SDG 4: Ensure inclusive and quality education for all and promote lifelong learning[14].
- ✓ In railway transportation operations, maintenance and safety[15].

At the Big Data Working Group (WGBD) meeting held in Copenhagen, Denmark in 2019, it was noted that compared with the last two working group meetings held in 2017 and 2018, all member SAIs attached great importance to the development of big data in audits and actively exploring the theoretical and practical applications, improving the development of big data audit platform tools, in-depth cooperation and exchange of big data auditing, and comprehensive capacity-building activities. Some applications of big data in the SAI audits presented were the following:

- ✓ Social and economic benefits by the GAO US.
- ✓ The UK National Audit Office shared the cases of "General Ledger Analysis Application", "BBC Network Analysis", "Accounting Policies Text Analysis", "Test of Detail Application" and "Public Facing Analytics".
- ✓ Turkey uses audit software to conduct big data auditing through scenario analysis. Cases of "Analysis on Public Officials' Salaries", "Analysis on Financial Statements" and "Risk Assessment on Municipalities' Data" were shared.
- ✓ "Data-based case selection", "Increase of highway speed limit", "Access to long-term care benefits" and "Data ethics at Tax and Custom office" by the Netherlands Court of Audit.

[9] Desarkar, A. & Das, A. 2017. Big-Data Analytics, Machine Learning Algorithms and Scalable/Parallel/Distributed Algorithms. *Internet of Things and Big Data Technologies for Next Generation Healthcare*, 159-197.

[10] Al Nuaimi, E., Al Neyadi, H., Mohamed, N. & Al-Jaroodi, J. 2015. Applications of big data to smart cities. *Internet Serv Appl*.

[11] Kropp, T. 2018. *HHS OIG Fraud Analytics.* Washington DC: HHS OIG.

[12] Office of the Comptroller and Auditor General of India. 2017. *Office of the Comptroller and Auditor General of India.* New Delhi: Office of the Comptroller and Auditor General of India.

[13] Mpatisha, B. 2019. Big Data: A Case of China. *AFROSAI-E Technical Update.* Cape Town: AFROSAI-E.

[14] Vehkasalo, V. 2018. *Auditing Sustainable Development Goals Using Data Analysis: Dropout Prevention Program in Finnish Vocational Education.* National Audit Office of Finland.

[15] Ghofrani, F., He, Q., Goverde, R. M. & Liu, X. 2018. Recent applications of big data analytics in railway transportation systems: A survey. *Transportation Research Part C: Emerging Technologies*, 226-246.

- ✓ The Estonian National Audit Office overcame the problem of missing information and fully exploited the value of information in the case of "What is the status of local governments' real estate, what services are provided there and is the network of service optimal?". Four steps, i.e. obtaining an overview of all buildings, getting the objects on the map, establishing where beneficiaries live, and determining where the EU funding has been allocated, are used to make the objective audit evaluation.
- ✓ SDG audit development in Thailand. Thailand is a member of the ASOSAI project group and leads the monitoring part in the research project "Audit on the implementation of SDGs: leverage digital or big data to achieve the SDGs"[16].

2.3 Factors affecting big data analysis

There are some fundamentals that need to be in place for SAIs to conduct big data analysis. Some requirements that have been identified[17] are the following:

- ✓ Identify opportunities for the use of big data in audit − the SAI must assess what value and benefits the use of big data will bring to audits.
- ✓ There must be an identified mission and vision, and strategic and operational objectives.
- ✓ Identify priorities that should be in place like financial or strategic plans to ensure that the appropriate investments are made.
- ✓ Big data management and governance.
- ✓ Big data processing platforms − big data applications need to perform data analytics that usually require huge processing capability.
- ✓ Adequate network infrastructure.
- ✓ Advanced algorithms − standard algorithms used in regular applications may not be sufficient or efficient enough to handle big data applications due to their unique requirements and pressing need for high-volume high-speed processing.
- ✓ Citizen awareness – specifically for smart cities, citizens must be aware of how to use ICT solutions and provide information.
- ✓ Government role − to establish guiding principles of openness, transparency, participation and collaboration to keep the exchange and flow of big data under control.[18]

Experiences shared at the INTOSAI Working Group on Big Data (WGBD) show that some SAIs using big data in their audits have worked or are working on the following fundamentals:

- ✓ Big data strategy (SAI Denmark).

[16] WGBD. 2019. Summary of Big Data Audit Experience. *3rd Meeting of the INTOSAI Working Group on Big Data.* Copenhagen: WGBD.

[17] Al Nuaimi, E., Al Neyadi, H., Mohamed, N. & Al-Jaroodi, J. 2015. Applications of big data to smart cities. *Internet Serv Appl*.

[18] Al Nuaimi, E., Al Neyadi, H., Mohamed, N. & Al-Jaroodi, J. 2015. Applications of big data to smart cities. *Internet Serv Appl*.

- ✓ The UK National Audit Office has introduced data analytics for audit at the NAO and a future model for data analytics workflow.
- ✓ The European Court of Auditors has established a high-level task force on strategy and future foresight for conducting big data audit activities to output multi-annual strategy and broad, systematic trend analysis, consider future developments and emerging issues, guide the selection of audit and review tasks and give input for strategic briefing products and external and internal events. The European Court of Auditors has set up an audit innovation laboratory that brings together data scientists, artificial intelligence experts and auditors interested in technology to conduct relevant scenario testing.
- ✓ The Netherlands Court of Audit has emphasised the construction of a big data audit team and the cultivation of data awareness and formed a complete data analysis team.
- ✓ Exploring the application of new technologies and methods in big data auditing.
- ✓ The development of big data audit-related platform tools.
- ✓ The State Audit Office of the Kingdom of Thailand introduced an application of big data and data analytics in auditing and a guideline on effective integration of big data into auditing.
- ✓ Big data audit cooperation and capacity building.[19]

2.2 Objectives of the theme study

The primary objective of this study is to establish the fundamentals needed for SAIs to integrate big data in public sector auditing. The theme paper, which is a case study of SAIs in AFROSAI-E and the CREFIAF region, seeks to answer some of the following questions:

- ✓ What are the levels of country computerisation and data generation?
- ✓ What data do countries keep – what are the data sources, structures and data-handling capabilities of countries?
- ✓ Is there an adequate understanding of big data – in audit and in government at large?
- ✓ Have the SAIs established roles and responsibilities and identified stakeholders?
- ✓ Do the SAIs have, or are they able to get, big data on which to apply advanced analytics?
- ✓ Do SAIs have the people, process and technology in place to build capabilities that will make productive use of data that the SAI has collected?
- ✓ Are there significant opportunities to utilise big data analytics in public audit and government in general?

[19] WGBD. 2019. Summary of Big Data Audit Experience. *3rd Meeting of the INTOSAI Working Group on Big Data.* Copenhagen: WGBD.

2.3 Significance and aims of the study

ISSA1 12 calls on SAIs to demonstrate ongoing relevance to citizens and be responsive to changing environments and emerging risks[20]. The increased and varied automation in the African public sector implies a changing audit landscape for SAIs. The tools that the SAIs currently have at their disposal may no longer be adequate to sort, verify and analyse the increasing amount of data compiled by governments. Further, in the spirit of ISSAI 12, SAIs must promote increased transparency in government. SAI Estonia has seen the role of a modern SAI as promoting the creation of a data exchange environment at government level and nudging the auditees to perform relevant data analysis[21]. Building bridges between authorities to enable data sharing can become a tool for advancing the country's decision-making and improving the wellbeing of its people.

Some studies on big data have focused on investigating the general requirements for implementing big data application. The AFROSAI-E research team is of the view that given the multiple identified uses, which in many cases are applicable to public audit, there should be more of a focus on the requirements for integrating big data into public sector audit. Big data analytics needs a system architecture for data collection, transmission, storage, processing and analysis, and visualisation mechanisms[22]. Other scholars[23] advocated creating a roadmap for success in using big data that would cover several stages. This study therefore focuses on the aspects of the mechanisms needed for integrating big data in public sector audit.

The study is aimed at SAIs and will firstly give an overall understanding of what big data is. Secondly, it will highlight the opportunities and potential challenges big data brings to public sector audit. In highlighting these challenges, the paper will lay a base for deciding how SAIs can proactively overcome these challenges. Further, the paper seeks to clarify some of the misconceptions about big data or the use of big data in audit.

The expected results of the study were:

- ✓ A research paper contributing to current literature for AFROSAI and the INTOSAI community at large to use on the subject of big data in public audit.
- ✓ A paper highlighting future considerations and/or guidance for big data in public sector auditing.

[20] ISSAI.Org. 2020, May 28. ISSAI.Org. Retrieved from ISSAI.Org: https://www.issai.org/wp-content/uploads/2019/08/INTOSAI-P-12.pdf

[21] Holm, J. 2020, February. 'Big data' analysis and modern supreme audit institutions: tearing down the walls of data kingdoms. National Audit Office of Estonia.

[22] Bendre, M. R. & Thool, V. R. 2016. Analytics, challenges and applications in big data environment: a survey. *Journal of Management Analytics*, 206-239.

[23] Al Nuaimi, E., Al Neyadi, H., Mohamed, N. & Al-Jaroodi, J. 2015. Applications of big data to smart cities. *Internet Serv Appl*.

- ✓ Guidance for the SAIs preparing country papers on the theme "Integrating Big Data in Public Sector Audit".

2.4 Timelines and project plan – June to August

In preparing the study, AFROSAI-E developed a detailed project plan indicating the administrative mechanisms, responsibilities and proposed dates. The table below shows the detailed activities and actual timelines of the study.

Table 1: Study project plan

Activity	Responsibility	Date
1. Assignment of theme to AFROSAI-E	AFROSAI	3 Feb 2020
2. Introduction to theme	AFROSAI-E CEO & EOs	Feb 2020
3. Context of theme and study proposal	AFROSAI-E IS Audit Managers	May 2020
4. Development of questionnaire	AFROSAI-E IS Audit Managers	30 May 2020
5. Testing of research tool (questionnaire)	SAIs – Zambia, Tanzania, Namibia	8 June 2020
6. Communication to SAIs and distribution of questionnaire	AFROSAI-E CEO	15 June 2020
7. Launch of study and online seminar	AFROSAI-E IS Audit Managers	15 June 2020
8. Interview with private audit firms	AFROSAI-E IS Audit Managers	19 June 2020
9. Translation of questionnaire into French	AFROSAI	29 June 2020
10. Administration of research instrument and field work	AFROSAI-E IS Audit Managers	7 July 2020
11. Data analysis and research writing	AFROSAI-E IS Audit Managers	17 July 2020
12. Reviews and finalisation	AFROSAI-E CEO, EOs, Technical Managers, partners	August 2020
13. Final paper	AFROSAI-E CEO	August 2020
14. Guidance for country papers	AFROSAI-E CEO	August 2020
15. Translation of paper into French	AFROSAI	August 2020

3 THEME STUDY METHODOLOGY

3.1 Introduction

This section provides a description of the research procedures adopted in this study. It explains the study design, the selection of study population and sampling techniques. It further describes the research instruments used in data collection and presents how the research team collected and analysed the data, with reference to the objectives of the study.

3.2 Research design

To address the research objectives more effectively, this research undertook a cross-sectional study and made use of both quantitative and qualitative approaches. The methodology is a combination of descriptive, exploratory and explanatory, hence the following combination of research strategy and the reasons behind the use of such strategy:

1. Survey – In line with Anderson[24], a survey method was used because of the nature of the research, which can be classified under the applied social sciences, and it involved the use of a questionnaire as a data collection technique.
2. Case study – A case study method was used because the research itself is an in-depth study of a regional grouping of SAIs in Africa called "AFROSAI". Accordingly, in evaluation, case studies can be used to capture the complexity of a case, including temporal changes, as well as exploring the contextual conditions of a case.[25] It can thus be argued that a case study provides more realistic responses than a purely statistical survey.

The researcher collected quantitative data from a sampled set of participants from SAIs in the AFROSAI region. The researcher wrote to the Heads of SAIs and gained permission for at least two SAI staff to take part in the surveys. The sample consists of SAI employees deemed by virtue of their managerial position within the organisation as the most knowledgeable individuals to assist towards meeting the research objective by responding objectively.

3.3 Sample size determination

The size of the sample is an important parameter of the sample design because it affects the precision, cost and duration of the survey more than any other factor. Its determination was based on some statistical determinants such as margin of error, design effect and total population. The Creative

[24] Anderson, R. Intuitive inquiry: An epistemology of the heart for scientific inquiry. *The Humanistic Psychology*, 32(4), pg. 307-241.
[25] Ibid.

Research Systems[26], a sample size calculator which makes use of the above three stated statistical determinants, was employed in the determination of a statistically appropriate sample size. The projected population of respondents was 98. Using the sample size calculator and allowing an error margin of 10% and 95% confidence level that the response from the sample was representative of the entire population, this population size gives the desired sample size of 49.

A total of 32 respondents from 22 SAIs (17 Anglophone and 5 Francophone) took part in the survey. Additionally, the research team utilised multiple data sources to get further assurance and credibility regarding data.

3.4 Data instruments

The data collected comprised secondary and primary data. Secondary data was collected through the review of published research documents, while primary data was collected by using both quantitative and qualitative approaches through structured and open-ended questions.

3.5 Data collection

The researcher collected data by using the questionnaire (Appendix 1) from 15 June to 7 July 2020. To ensure that the data collected was of high quality, several quality assurance mechanisms were implemented. The researchers conducted a webinar to explain the study objectives and clarify any issues regarding the study instrument or the study. After data collection was completed, the researcher edited the data prior to the data entry and processing. Before the analysis was performed, data cleansing was done to ensure consistency and completeness.

[26] https://www.surveysystem.com/sscalc.htm#one <Online accessed 03 August 2020>

4 THEME STUDY RESULTS

4.1 General information

We received 32 responses in total, from both CREFIAF (22%) and AFROSAI-E (78%) member SAIs. The received responses cover the regions of Africa as shown in figure 4-1 below.

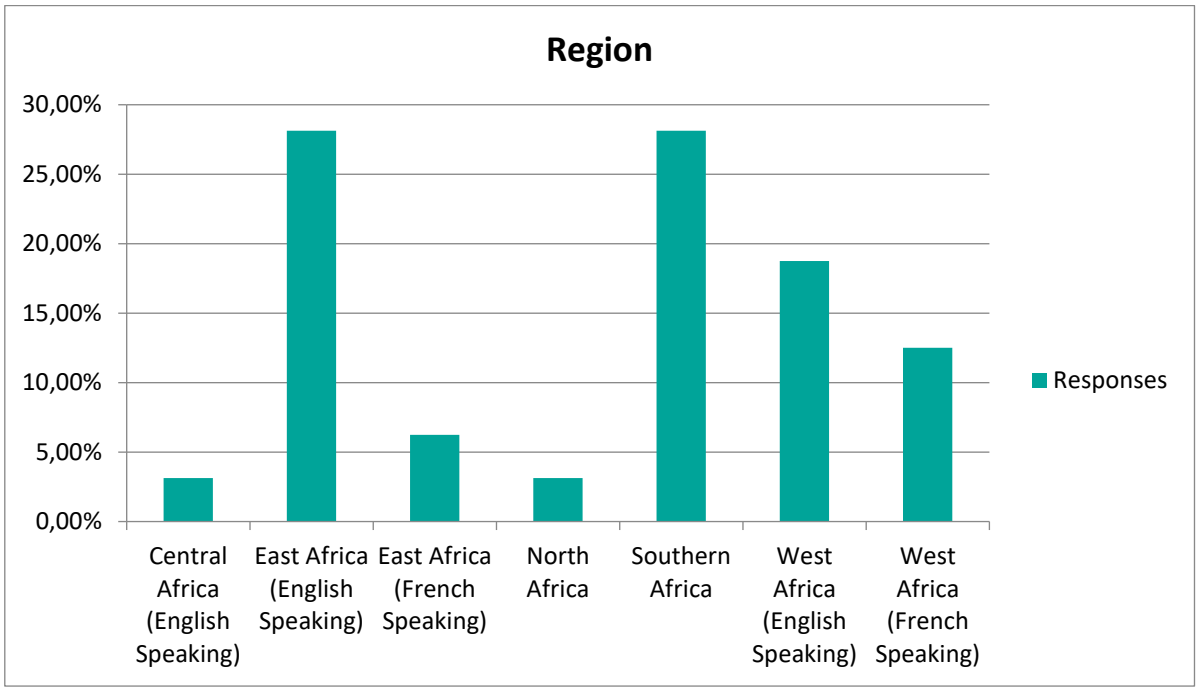

Figure 4-1 Respondents per region

This study on integrating big data in public audit is a technical theme paper on how SAIs can apply analysis of big data in audit. The technical and practical audit knowledge of the respondent is therefore key. The study encouraged SAIs to have at least two participants with the intention that one should have a technical background, while the other had a managerial background. A few general questions gave the option of responding from a management perspective. The more technical the respondent is, the more likely they are to be knowledgeable about the subject matter of big data.

Approximately 45% of respondents were auditors, 10% were in audit management and 15% were in information system management (Figure 4-2), while about 29% were on the SAIs' executive management level. The composition of the respondents reassured the researcher as to knowledge of the big data subject matter.

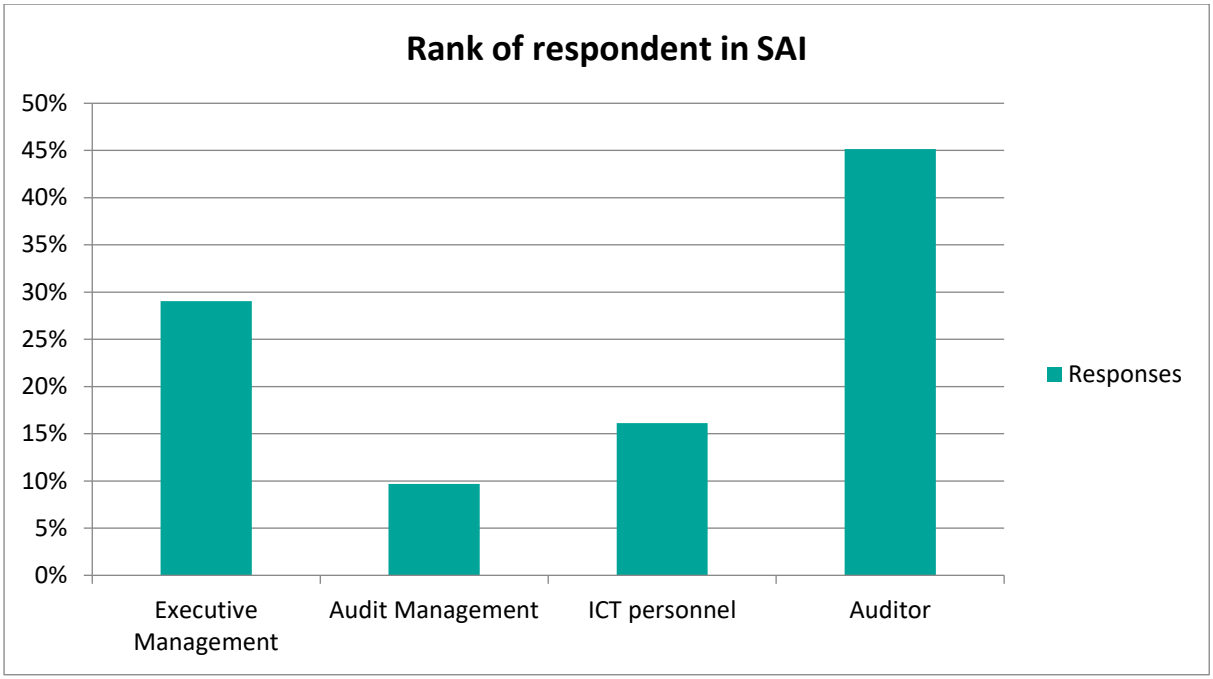

Figure 4-2 Rank of respondent in SAI

Table 2 below shows that a majority of the respondents, approximately 58%, had more than ten years of experience in the SAI. About 22% of the respondents had one to five years of experience while 19% had six to ten years of audit experience in the SAI. The study drew information from a range of respondents with good experience in the SAI, which enabled the research team to gain relevant insights on the use of big data in the public sector.

Table 2 Respondents' years of experience in the SAI

Respondents' years of experience in the SAI	Count
1 – 5	7
6 – 10	6
More than 10	18
	31

In its framework of pronouncements, INTOSAI recognises three types of audits (performance, financial and compliance). In practice, practitioners tend to classify IT and special audits as stand-alone audits[27].

Most respondents, approximately 53%, were from IS Audit and IS management functions within the SAI. 15% were drawn from Financial Audit, 6% from Performance Audit and 25% from Executive

[27] AFROSAI-E. 2020. *Resilience of SAIs during COVID19.* Pretoria: AFROSAI-E.

Management. Most of the respondents were assumed to have a good understanding of the principles and uses of big data. See figure 4-3 below.

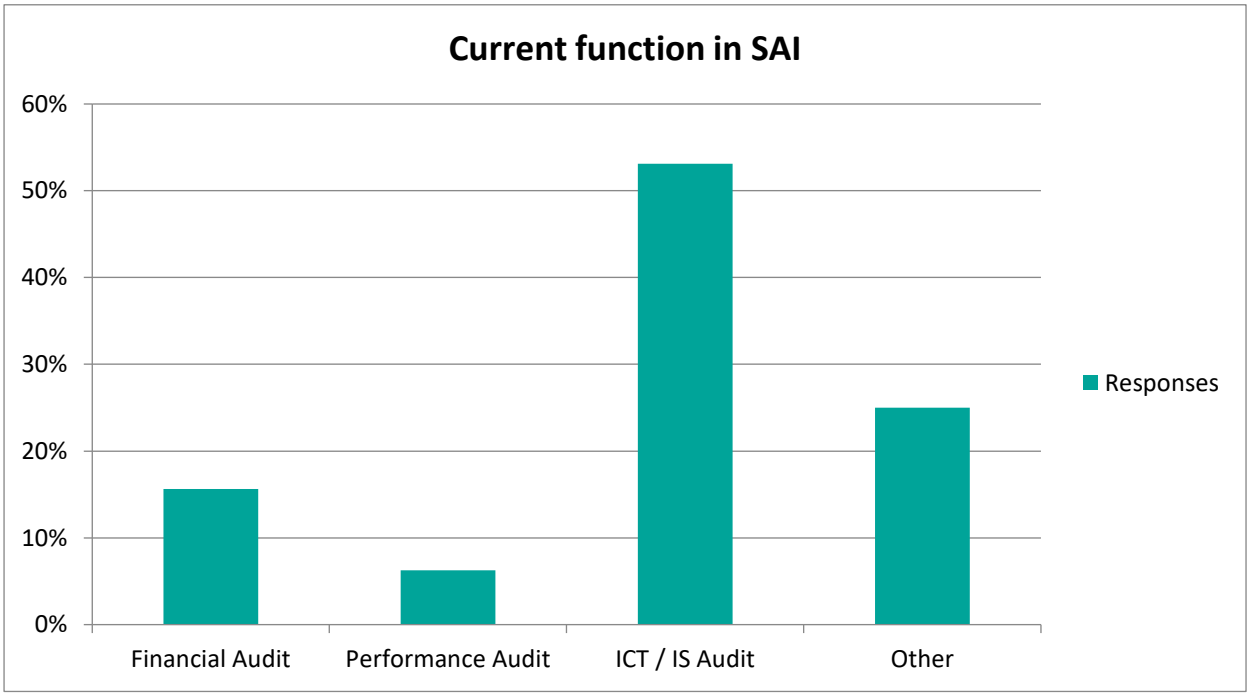

Figure 4-3 Respondents' current function in the SAI

4.2 Understanding data analytics and big data

Although data analysis has been part of audit for decades, it has only recently become standard practice. As most, if not all, organisations are impacted by IT, it is essential to use technology to conduct an effective audit. SAIs may use technology to improve the audit and reduce the time necessary to complete the engagement. The audit standards already require the consideration to use data analysis. Auditors can apply data analysis in all phases of an audit, in addition to identifying errors or fraudulent activities. Data analysis does not change the objective or scope of an audit; it is simply another technique applied to achieve the objective of the audit.

According to INTOSAI[28], data analysis is a process of inspecting, cleansing, transforming and modelling data with the goal of discovering useful information, informing conclusions and supporting decision-making.

Data analysis is most effective when implemented using data analysis technology, which enables auditors to manipulate a complete data set − 100% of the transactions in a population − to analyse

[28] INTOSAI Working Group on IT Audit. 2019. *Data Analysis Guideline.* Vienna: INTOSAI.

the relationships and correlations between data sets, identify patterns and trends, and visualise results graphically[29].

Big data analysis can simply be termed as conducting data analytics on big data, counted in hundreds of gigabytes, terabytes or even petabytes, with a frequent and rapid growth and change, which in some cases can be in real time[30]. SAIs should keep this context in mind when assessing needs and possibilities for big data analytics. Big data is data with sizes and types that exceed the capacity of traditional software to process within an acceptable time and value. In conducting big data analytics, the traditional tools of analysis are also inadequate to capture, manage and process the data.

Big data analytics has been described as the process of examining information and patterns from large data. To be able to do this, the person carrying out the analysis needs a system architecture for data collection, transmission, storage, processing and analysis, and visualisation mechanisms[31].

4.2.1 Auditors' knowledge and use of data analysis

Before researching the implementation and application of big data, SAIs have to build capacity in data analysis. SAIs should be able to analyse larger quantities of data from different sources even if it is mostly structured data, and to deal with data that is not volatile, i.e. looking at data at a fixed point in time. The SAIs should also be able to use different methods and tools for data analysis to tackle the challenge of larger quantities of data. These basic aspects are important to handle before putting big data on the agenda.

Hence, we asked the respondents to give an indication of their understanding of data analytics. Approximately 41% of the respondents had a good understanding of data analytics and 9% had a very good understanding. The other 48% had a fair or not good enough knowledge of data analytics. See figure 4-4 below.

[29] ICAEW. 2016: Data analytics for external auditors.
[30] ISACA. 2013, March. Big Data Impacts and Benefits. An ISAXA white paper. Isaca.org
[31] Bendre, M. R. & Thool, V. R. 2016. Analytics, challenges and applications in big data environment: a survey. *Journal of Management Analytics*, 206-239.

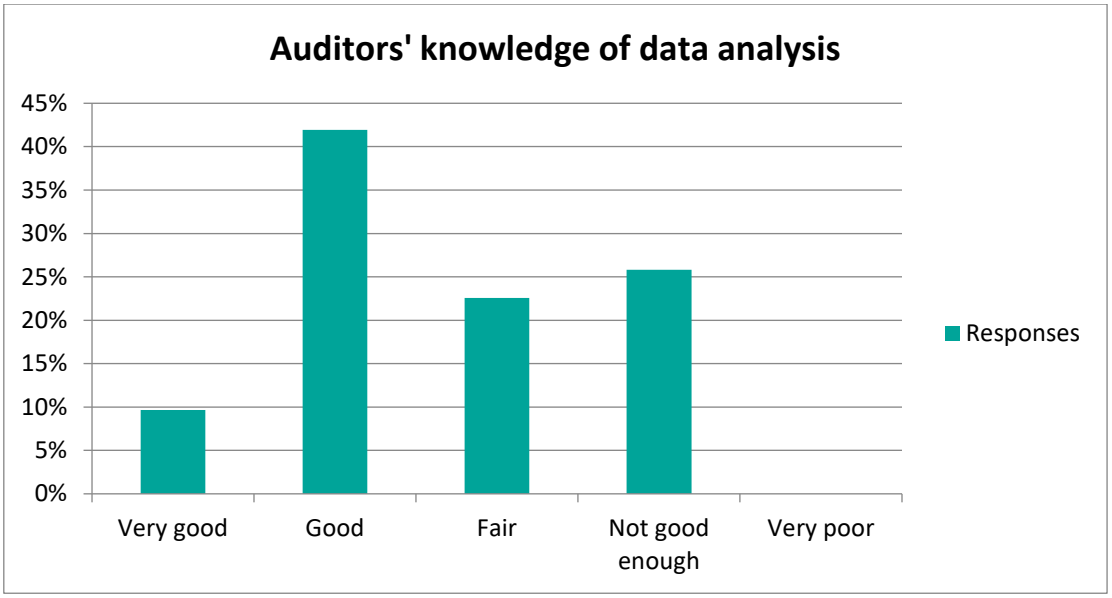

Figure 4-4 Auditors' knowledge of data analysis

This observation is important because it shows that African SAIs still have a way to go in building capacity in data analytics and ensuring coverage in all types of audit.

Most of the respondents, however, indicated that they were comfortable with conducting data analysis as part of their audit work (70%). Approximately 29% were not sure or uncomfortable with the use of data analysis in their audit. See figure 4-5 below.

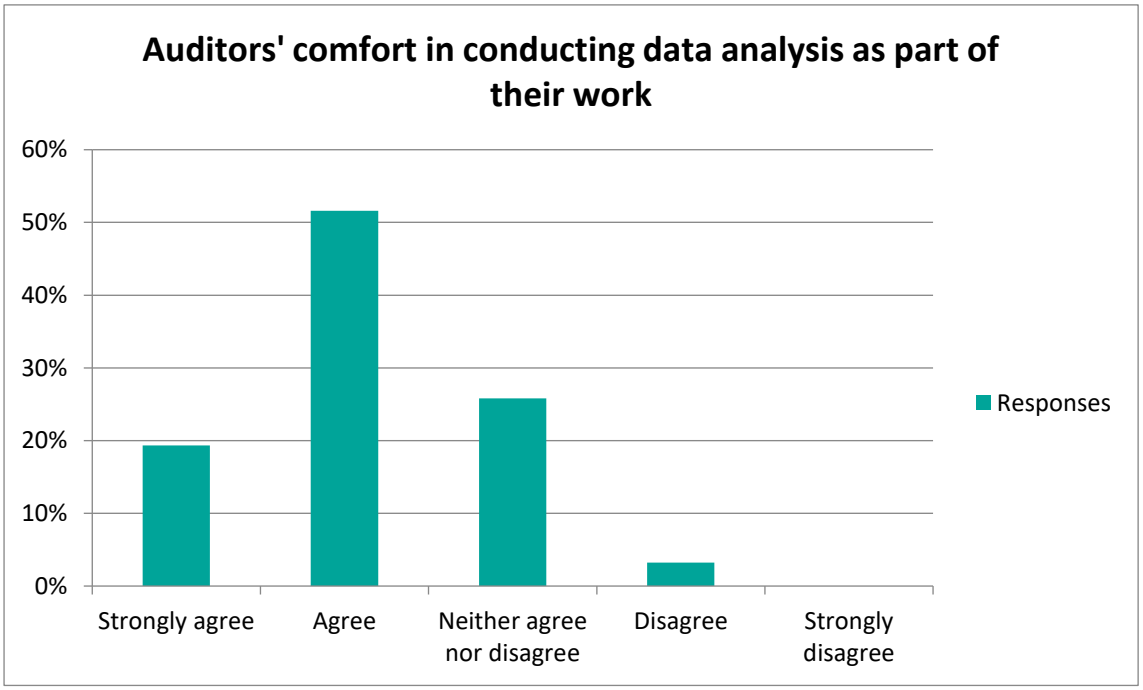

Figure 4-5 Auditors' comfort in conducting data analysis as part of their work

This could be due to the composition of the responders, many of whom worked in the IT audit units that are often assigned the duties of data analytics. Nevertheless, it does not account for the gap between knowledge of data analysis and integrating it in one's work. Most auditors will conduct standard analysis every year and hence become very comfortable with that approach. However, when it comes to expanding the analysis to cross-sectional investigations or merging data sets, a lack of knowledge may limit the SAIs in their work.

4.2.2 Knowledge of big data

The concept of big data is somewhat unclear to many auditors with the common misconception being that any large amount of data qualifies as big data. Hence, the SAI respondents were given the characteristics of big data defined in section 2.1, together with the study brief.

The study sought to gain an indication of the respondents' understanding of big data terminology. Approximately 45% of the respondents had clarity on the big data terminology (13% strong clarity and 32% good clarity). 39% of the respondents were indifferent as to their understanding and the other 16% were not clear about big data terminologies. See figure 4-6 below.

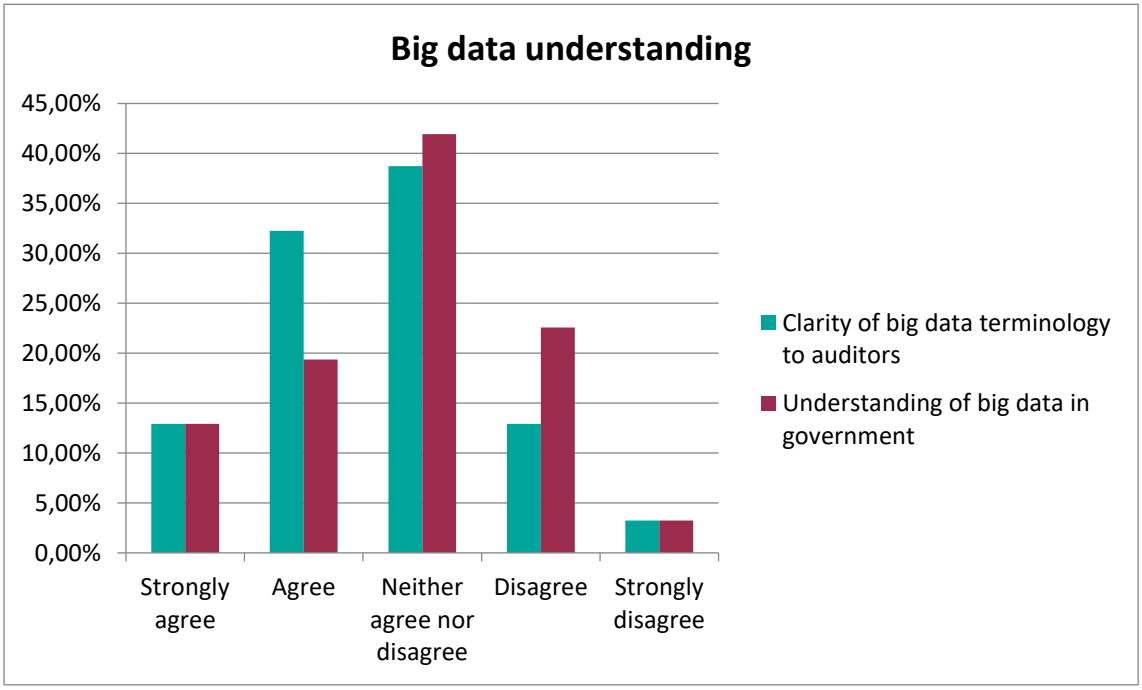

Figure 4-6 Big data understanding – government and public auditors

The implementation of technology, processes and capabilities for big data analytics is a huge undertaking, which will involve large parts of the organisation. This result shows a lack of maturity in the participating SAIs that they have to address before making any investment.

The respondents' sentiments on the understanding of big data in government were worse than their opinion of the auditors' clarity on big data. Only 32% of SAI respondents either agreed or strongly agreed that there generally was an adequate understanding of big data in government. 41% of the respondents neither agreed nor disagreed while 26% either disagreed or strongly disagreed. Evidently, there is an awareness gap regarding big data. See figure 4-6 above.

4.3 Digitalisation and data handling in African countries

There are various statistics on data handling and levels of computerisation of African countries. This computerisation, which is increasing[32], has not seen a corresponding change in legal frameworks. In a study by the World Economic Forum conducted in 139 countries, there were no African countries in the top 40 with laws relating to the use of information and communications technologies (ICTs) comparable to those of developed countries[33]. The report however highlighted the importance of technology to the African countries with several African countries appearing in the top 40 list of government procurement of advanced technology products and importance of ICTs to governments' vision of the future. Rwanda, for instance, was number 6 on government procurement of advanced technology products and number 4 on importance of ICTs to government's vision of the future.

The National Audit Office of the People's Republic of China (NAO) conducted a survey in INTOSAI on how various governments made use of IT in public service delivery. According to GIZ (2019) some of the findings were:

- ✓ 45 out of 64 respondents (70%) reported that their country had adopted IT systems for public administration in most departments.
- ✓ 40 out of 64 respondents (63%) indicated that most government departments provided public services via internet.
- ✓ 33 out of 64 respondents (52%) reported that departments shared data with other departments[34].

4.3.1 Country digitalisation

Digitisation is the process of converting information from a physical format into a digital one (i.e. computer-readable)[35]. When digitisation is leveraged to improve business processes, it is called digitalisation. The results of this process, like providing better public services, are called digital transformation. Technology and digitalisation are driving many aspects of lives, e.g. digital cities, digital banks, digital citizens, e-government. These technologies, previously few in Africa, are currently on the increase and being managed on the African continent. Smart Africa, an African Union alliance

[32] Internetworldstats. 2020, July 15. Internetworldstats. https://www.internetworldstats.com/stats.htm
[33] World Economic Forum. 2016. *The Global Information Technology Report 2016*. Geneva: World Economic Forum.
[34] Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). 2019. *Harnessing IT by SAIs: A Framework for Performance Improvement*. Germany.
[35] Schaefer, C. 2018. *The Future of Audit.* Pretoria: AFROSAI-E.

seeking to advance the usage of ICTs with 30 African countries, currently has the following, amongst others, flagship projects:

- ✓ High Tech Parks led by Angola
- ✓ Innovation and knowledge cities led by Benin
- ✓ African submarine fibre led by Cape Verde
- ✓ Digital economy led by Kenya
- ✓ Smart cities led by Rwanda
- ✓ Smart villages led by Niger
- ✓ Big data and data measurement for development led by Uganda[36].

Over the last decade, multiple uses of big data have been advanced or researched[37]. These studies have shown uses in areas where African governments have digitised or are working to digitalise. The uses of big data by countries or entities depend on the level of automation, data creation, data storage and availability of tools. The more automated or digitalised, the more data will be created and be available for analysis.

We wanted to find out more about firstly, how the participants rated their countries' progress in digitalisation and secondly, whether the countries' key processes were automated (e.g. financial reporting, budgeting, and key ministry roles, like citizen registration, passport systems, municipality accounts, vehicle registrations, education, health care etc.).

More than two thirds of the respondents rated their country's computerisation and digitalisation to be between fair and very good, with 13% rating their country's computerisation as very good, 30% as good and 23% as fair, whereas 27% indicated that it was not good enough, while 7% rated it as very poor. See figure 4-7 below.

These results are contradicted by numerous international indices. The Network Readiness Index (NRI)[38] is a global index of the application and utilisation of inform. High-income countries in Europe and North America dominate this ranking, whereas African nations dominate the lower ranks. There is a significant gap even between the top three in the region, with Mauritius ranked 53rd in the world, South Africa ranked 72nd and Rwanda ranked 89th.

A possible explanation for the contrasting views could be how the question was posed in the survey, not specifying whether it was meant globally or regionally. There are regional technology and

[36]Smart Africa. 2020, July 17. *Smart Africa*. Retrieved from Smart Africa. https://smartafrica.org/country-flagships/

[37] Al Nuaimi, Al Neyadi, Mohamed & Al-Jaroodi (2015), Bendre & Thool (2016), Ghofrani, He, Goverde & Liu (2018), Vehkasalo (2018), Schaefer (2018).

[38] The Network Readiness Index. 2019. https://networkreadinessindex.org/wp-content/uploads/2020/03/The-Network-Readiness-Index-2019-New-version-March-2020.pdf

digitalisation leaders in Africa, and as the NRI shows, there are significant differences among the African nations.

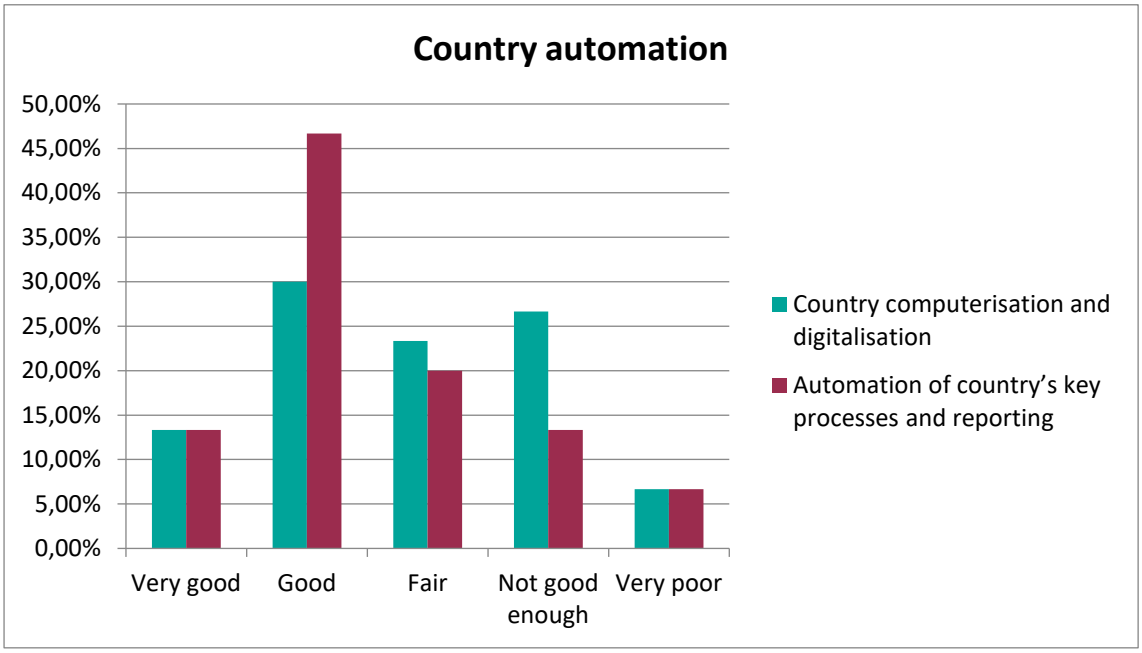

Figure 4-7 Country automation

All of 60% of the respondents agreed that key country processes were automated (13% strongly agreed and 47% agreed). 20% of the respondents neither agreed nor disagreed, 13% disagreed and 7% strongly disagreed.

However, the respondents' comments suggested their impression of this level of computerisation was because they had central systems for payroll and finance (IFMIS) and the governments had agencies that were implementing various e-Government projects. This may suggest that their judgement was based on limited knowledge of what an automated process should look like. Many key systems in Africa are poorly designed and not automated at all, which adversely affects basic service delivery.
The respondents who scored the level of computerisation as low, indicated that their country was still in an implementation phase. Another comment gave an example of government not having the right level of computerisation to enable remote work; hence, with the Covid-19 pandemic public auditors (and government) were not able to work efficiently remotely.

4.3.2 Data governance

Data governance is the foundation of a data-driven public sector. Elements of data governance are leadership and vision, coherent implementation, rules and guidelines, data infrastructure and architecture. Quality, relevant and timely data is essential to drive policies and programmes, whether it be in the efforts to create decent work for all, monitoring environmental degradation, containing

spread of virus or improving the living conditions in the urban areas[39]. To achieve goals and targets, like the SDGs, governments need to know if they are on the right path, hence they need data to monitor progress. Enabling the right cultural, policy, legal, regulatory, institutional, organisational and technical environment is necessary to control, manage, share, protect, and extract value from data. Still, public sector organisations often fail to govern, manage and value data in the same way as other assets that are critical to their success[40].

When asked how advanced their country's systems were for maintaining databases and registries for critical public data, only 3% rated them as very good and 30% gauged them to be good. Another 30% said the country's systems were fair, while 27% and 10% rated the systems as not good enough and very poor respectively. See figure 4-8 below.

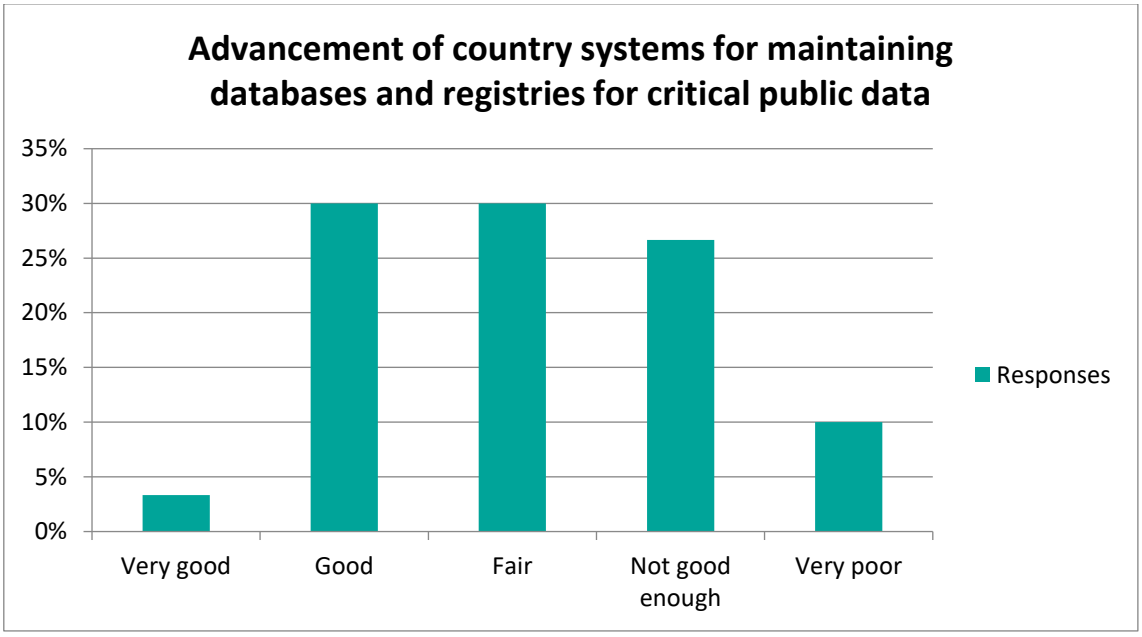

Figure 4-8 Advancement of country systems for maintaining databases and registries for critical public data

Good data governance can be very beneficial and is imperative for governments that aim to become more data driven as part of their digital strategy. It can help extract value from data assets, like social impacts, and put the country on the right track to achieving the SDGs. Further, enabling greater data access promotes transparency and accountability. However, many governments struggle to implement data governance frameworks and put it into practice.

23% and 40% of the respondents strongly agreed and agreed respectively that the public sector recognised data as a key strategic asset with its value defined and its impact measured. 23% were not sure and 13% disagreed that data was considered a strategic asset.

[39] Mohamed, A. 2020, May. 5th International conference on Big Data, Kigali. https://unstats.un.org/bigdata/blog/2019/planet-google.cshtml
[40] OECD Digital Government Studies. 2019: The Path to Becoming a Data-Driven Public Sector

7% and 60% of the respondents strongly agreed and agreed respectively that the public sector made active efforts to remove barriers to managing, sharing and re-using data. 20% were not sure and 13% disagreed that there were efforts to remove barriers to managing, sharing and re-using data.

7% and 40% of the respondents strongly agreed and agreed respectively that the public sector applied data to transform the design, delivery and monitoring of public policies and services. 43% were not sure and 10% disagreed that the public sector applied data to transform the design, delivery and monitoring of public policies and services.

7% and 60% of the respondents strongly agreed and agreed respectively that the public sector valued efforts to publish data openly as much as the use of data between, and within, public sector organisations. 20% were not sure, 10% disagreed and 3% strongly disagreed about efforts to publish data openly. See figure 4-9 below.

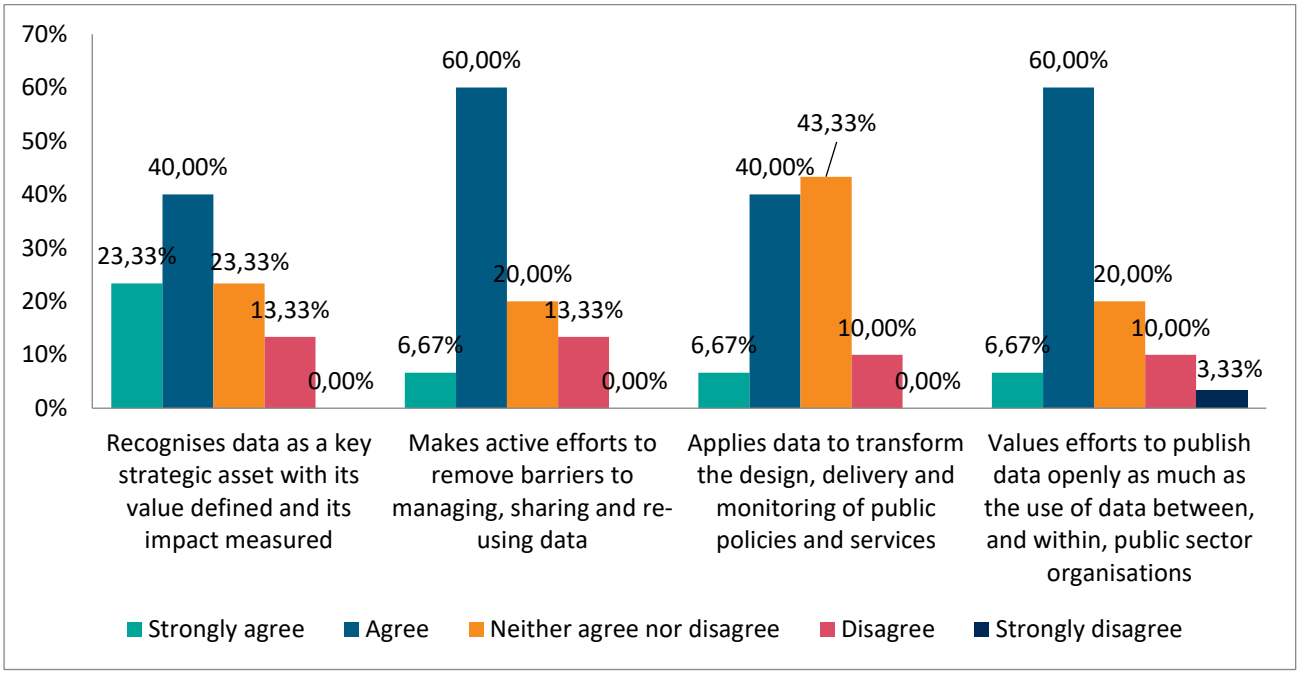

Figure 4-9 Views of data usage by the public sector

The Open Data Barometer (2016) measured governments' efforts to publish and use open data for accountability, innovation and social impact. Out of a top score of 100, the highest-ranking African countries, Kenya and South Africa, scored 40 and 34 respectively[41], whereas most African countries scored below 30. What affected the ratings of the African countries in this barometer negatively was that data was not available in machine-readable format and it was not re-usable, the data was not free of charge and not updated regularly. Most of the African countries also received a low score on

[41] World Wide Web Foundation. 2020, July 31. *Open Data Barometer*. Retrieved from Open Data Barometer: Country Automation.

positive impacts that open data release had had on a variety of different domains in the country. Kenya and Nigeria were exceptions here, scoring 57 and 41 respectively. Our study shows that while there is an intent and will in Africa, this is yet to be implemented.

To find out whether the governments' open data practices differed between those for the public and those across ministries, departments and agencies (MDAs), we asked the respondents specifically about the openness and availability of data for the public and for MDAs. It seems that government data is less available to the public than across MDAs, see figure 4-10 below. 7% of the respondents strongly agreed that government encouraged open data practices and government data was easily accessible across MDAs, compared to none when it came to accessibility for the public. For both questions, 30% of respondents agreed, while 37% (MDAs) and 47% (public) were not sure. 27% disagreed for MDAs versus 13% for the public. None of the respondents strongly disagreed that data was easily accessible across MDAs, compared to 10% for the public.

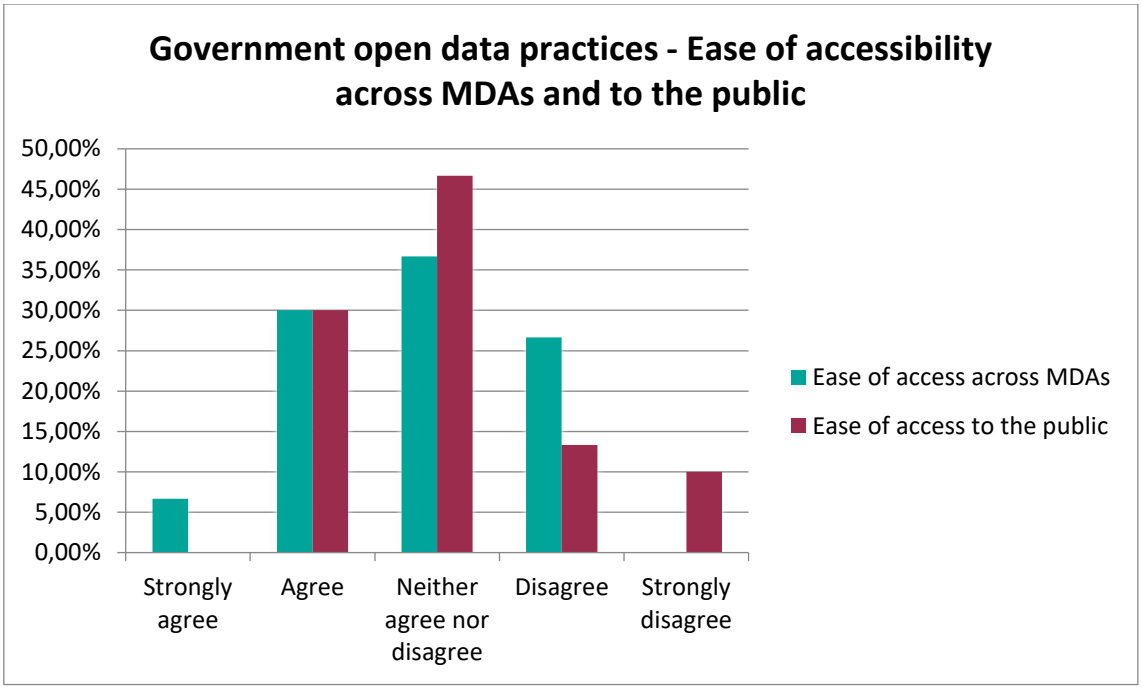

Figure 4-10 Government open data practices – ease of access across MDAs and to the public

4.3.3 Big data application in government

The value of data and information lies in the usage of it. Having systems to collect big data is only meaningful if the information is used to better decision-making, service delivery or product development. The AFROSAI-E research team asked respondents about the application of big data in the public sector and the opportunities.

Of the total responses received, 7% strongly agreed and 53% agreed that there was an understanding of potential uses of big data in the public sector. 17% of the respondents neither agreed nor disagreed

while 20% disagreed and 3% strongly disagreed about the understanding of uses of big data by government. See figure 4-11 below.

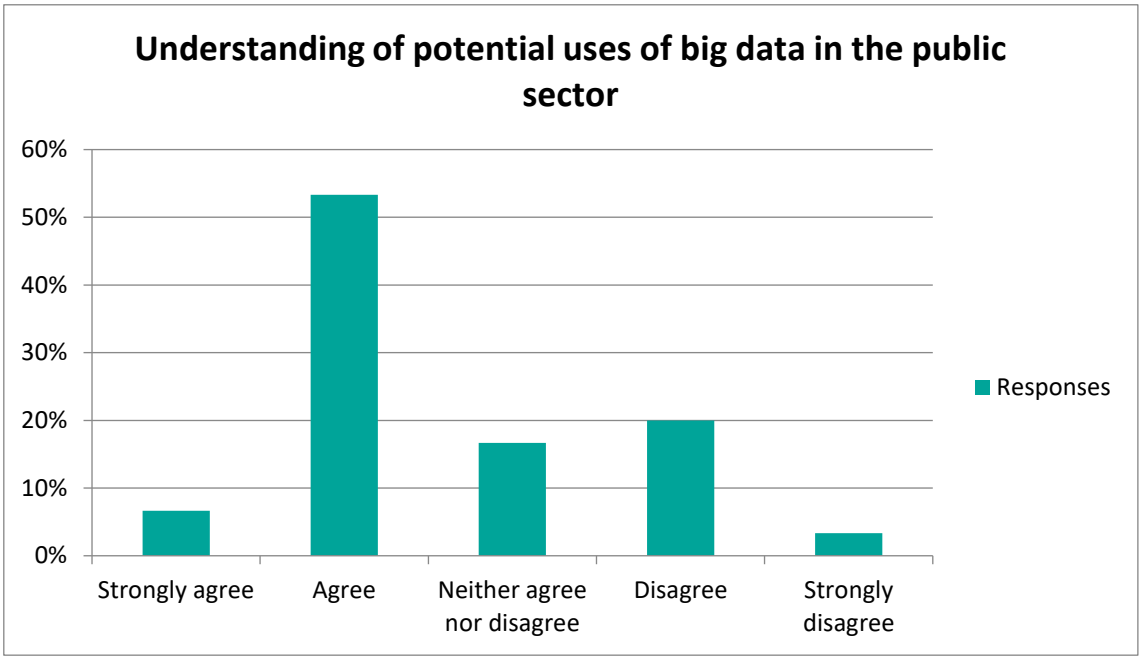

Figure 4-11 Understanding of potential uses of big data in the public sector

On the question of significant opportunities to utilise big data analytics in government services, 23% of respondents strongly agreed on the availability of opportunities and 43% agreed. The remaining 33% of respondents were indifferent or disagreed. The areas of use of big data included provision of services to citizens, fraud detection and security. See figure 4-12 below for responses.

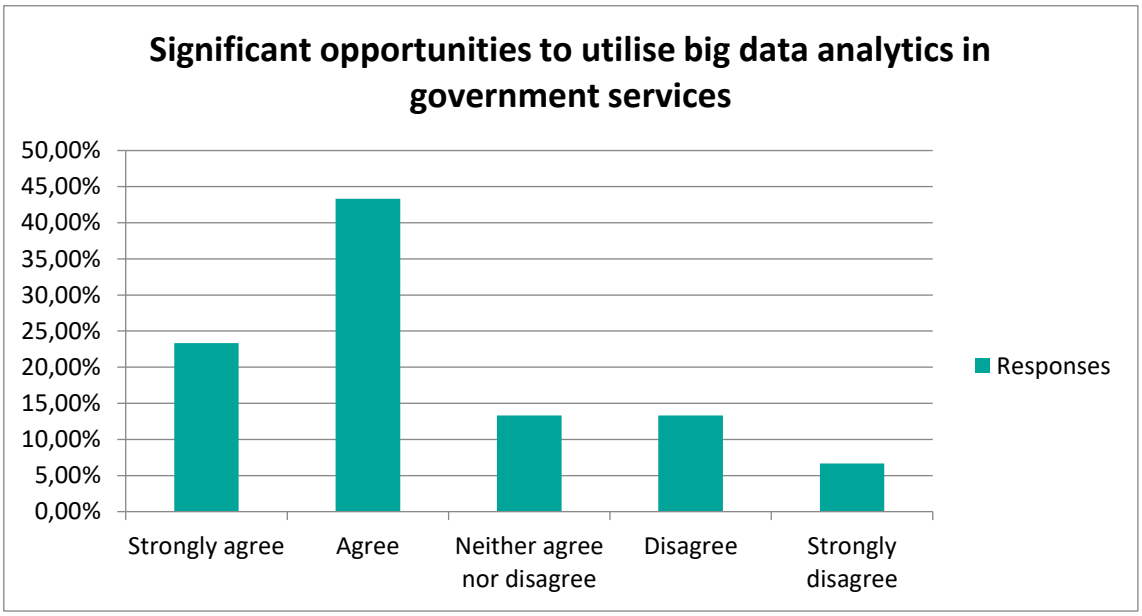

Figure 4-12 Significant opportunities to utilise big data analytics in government services

The comments given by the respondents on the opportunities to utilise big data analytics (appendix 4) were very general and did not indicate that any of the SAIs were aware of specific plans that their governments were implementing, nor that the SAIs followed up on these matters through audit.

There is a consensus that big data can be beneficial to governments to make better-informed decisions and to set and reach goals. We asked the respondents how their governments had used these opportunities. A majority of the respondents reported that their government either had not utilised big data in the provision of services to citizens or had utilised big data in very few ways (40% and 17% respectively), whereas 27% of respondents claimed their governments had used big data in the provision of services to citizens in several ways. Note that none of the respondents indicated that their governments had used big data in innovative ways.

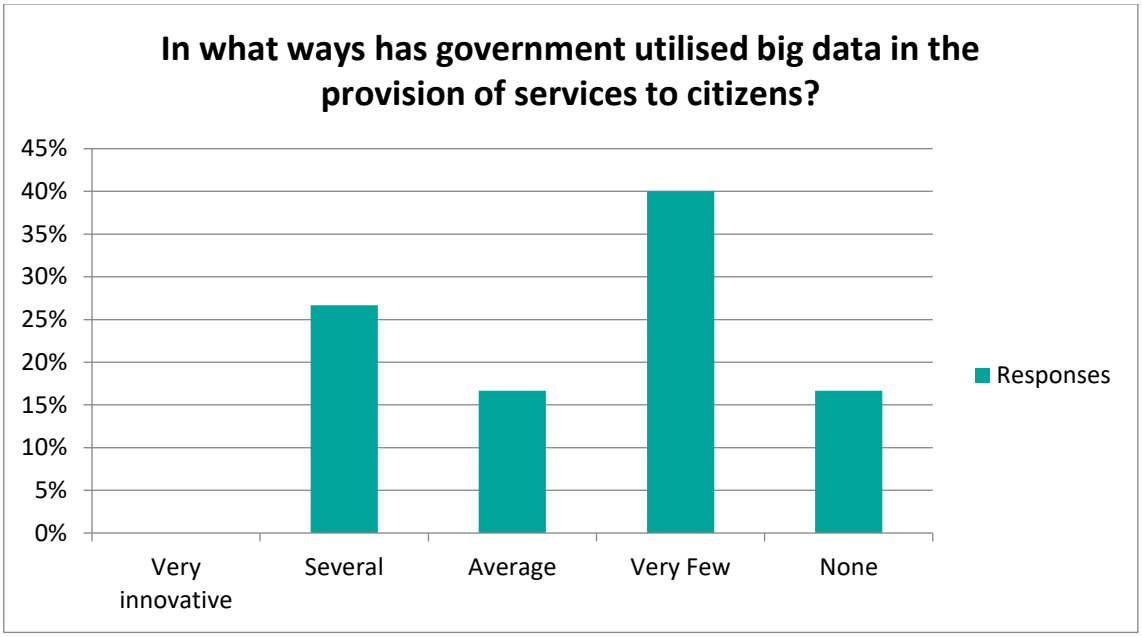

Figure 4-13 Government utilisation of big data in the provision of services to citizens

The examples given by the respondents were few and were not relevant to big data in service delivery, see appendix 4.

4.4 SAIs' use of data and big data analytics in audit

New analytics tools and methods are expanding the possible ways in which SAIs can derive value from existing data from government entities and from commercial data sources. While traditional data analytics has targeted structured data, like general ledgers, that can be easily analysed, advances in analytics methods now allow examination of more varied data types[42]. These advances could allow SAIs to attempt a wider variety of audit subject matters, issue better reports and make better recommendations.

[42] ISACA White Paper. January 2014. *Generating Value from Big Data Analytics.*

Big data analytics can also introduce additional technical and operational risks. The SAIs should be wary of the potential reputational risks that can be occasioned by analysing incorrect data or not using data analysis at all. SAIs need to ask the following questions to address key potential challenges before they can realise the gains from big data analytics:

- ✓ Does the SAI have the people, process and technology in place to build capabilities that will make productive use of data that the SAI has collected?
- ✓ Has the SAI established roles and responsibilities and identified stakeholders?
- ✓ Does the SAI have, or can it get, data on which to apply advanced analytics?

Not every SAI will be equipped to make use of big data analytic techniques. SAIs may fail to know even the basics: what big data actually is, what its benefits are and what infrastructure is needed. Many SAIs may be missing key skills in their existing personnel, or the technological infrastructure. In addition, the technical infrastructure may not be implemented in a way that allows access to data to make use of the information they collect. These questions are important because without a clear understanding, any investments made before readiness is fully achieved may be inefficient, produce suboptimal results, or represent needless expense.

The study sought to assess how big data could be integrated into public sector audit. Therefore, the largest focus in the main research instrument was on how SAIs were applying big data analytics. This core evaluation looked at the SAIs' (big) data universe, application of data analytics in audit, SAIs' capabilities to support big data analytics and applications of big data in public audit.

4.4.1 Data universe

To enable the use of (big) data in public audit, the SAIs need to know the data within their audit universe: who is collecting, generating, storing and processing it, and how easily the data can be accessed. SAIs have identified their audit clients well but in most cases do not maintain databases of information systems. This has been observed in the AFROSAI-E Institutional Capacity Building Framework (ICBF) assessments of percentage audit coverage of national information systems.

On the issue of whether the SAI maintained a database of key government information systems, 37% of the respondents agreed. 27% neither agreed nor disagreed while 23% and 13% of respondents disagreed and strongly disagreed, respectively. See figure 4-14 below.

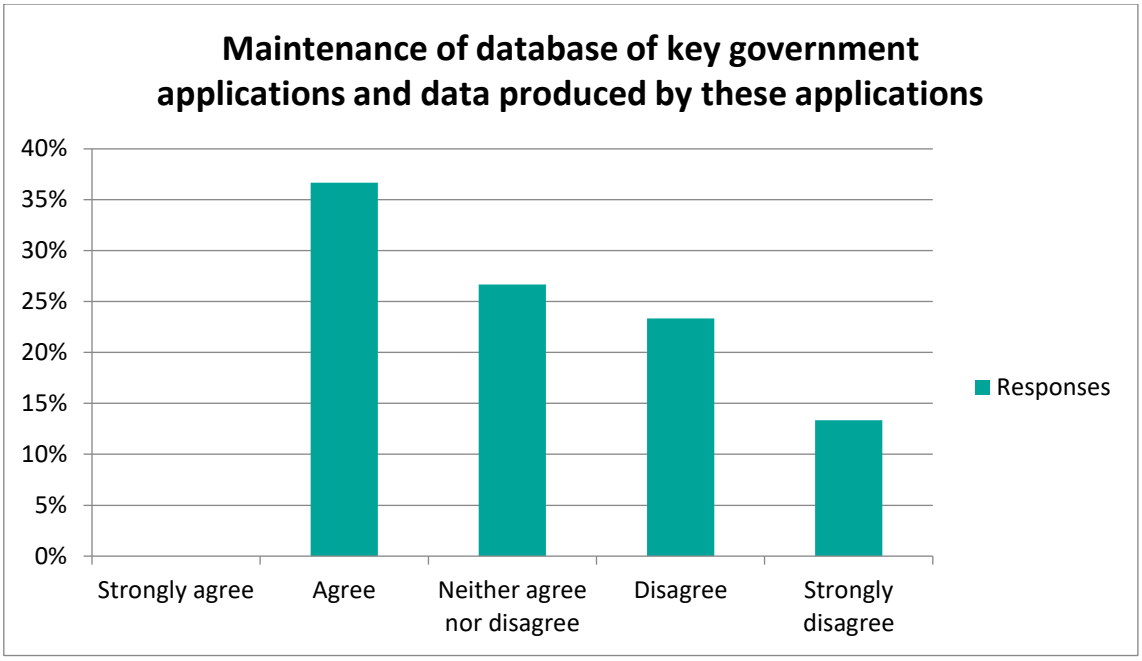

Figure 4-14 Maintenance of database of key government applications and data produced by these applications

Data flows have increased across organisations, sectors and borders. Data governance is no longer a matter limited to organisational boundaries. Identifying key stakeholders is important both in relation to access to data and to quality and completeness of data. The survey indicates that SAIs have identified stakeholders who are producing and processing data in government. From the survey, a majority of the respondents confirm this: 7% of respondents strongly agreed and 50% agreed that the stakeholders producing and processing information had been identified. 20% neither agreed nor disagreed while 17% and 7% of respondents disagreed and strongly disagreed, respectively. See figure 4-15 below.

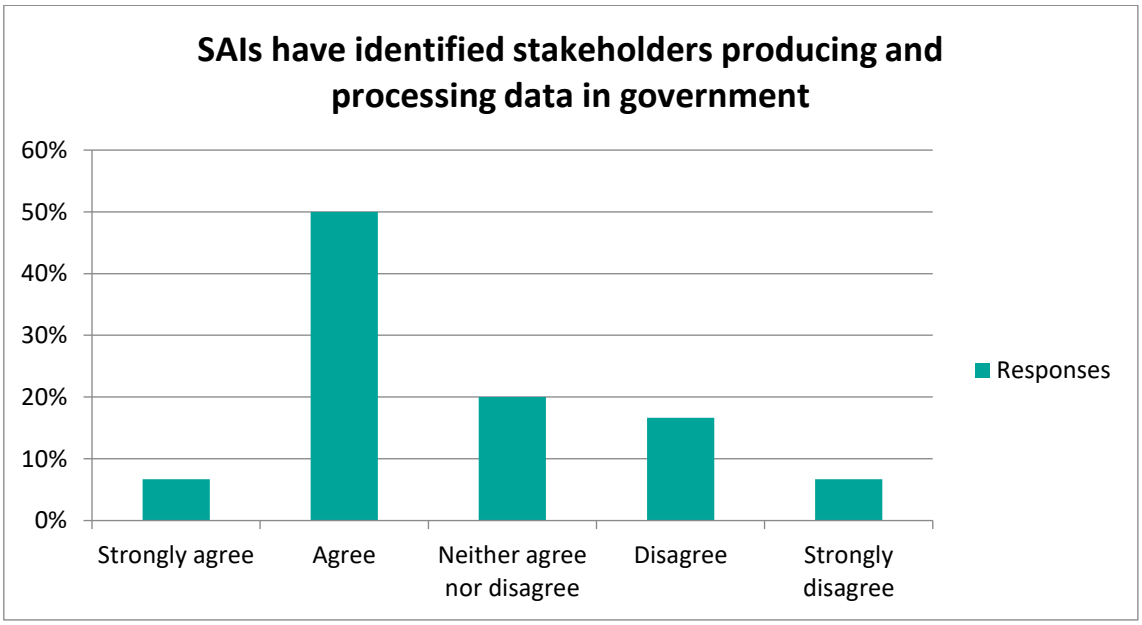

Figure 4-15 SAIs' identification of stakeholders producing and processing data in government

There is a discrepancy between the responses to data universe and identifying stakeholders. Although the SAIs should form partnerships beyond the public sector with regard to obtaining relevant data, the main data sources for a SAI will be their auditees. Whilst over half of the SAIs have identified stakeholders, it seems that they have not taken the appropriate steps to form partnerships to ensure access to relevant data.

The majority of respondents, however, agreed that government data was easily accessible to the public auditors: 10% strongly agreed and 60% agreed, whilst 13% neither agreed nor disagreed and 17% disagreed. See figure 4-16 below.

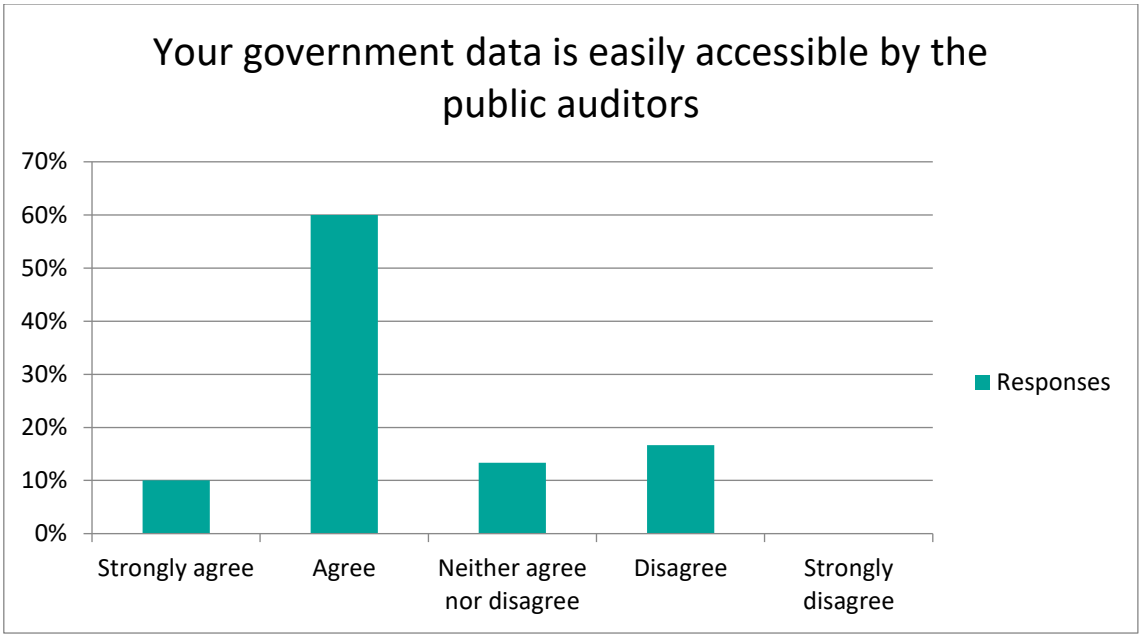

Figure 4-16 Ease of access to government data by the public auditors

The SAIs certainly have the legal instruments to ensure access to government data; however, they also have to ensure that the government entities comply with the laws and that they accept the results that emerge from analysing the data.

4.4.2 Data analytics in audit

Looking at how SAIs have implemented data analytics in their audits can give us an indication of skill sets in the SAI, like data literacy and analytic skills, and data governance. Building capacity in this area will be beneficial if the SAI has big data in its plans. We wanted to survey the extent of data analytics coverage to gain insight into existing data practices.

On the question of whether the use of data analysis was mandatory on financial audit assignments, 17% of the respondents strongly agreed and 27% agreed, while 33% of the respondents neither agreed nor disagreed, whereas 23% of the respondents disagreed. See figure 4-17 below.

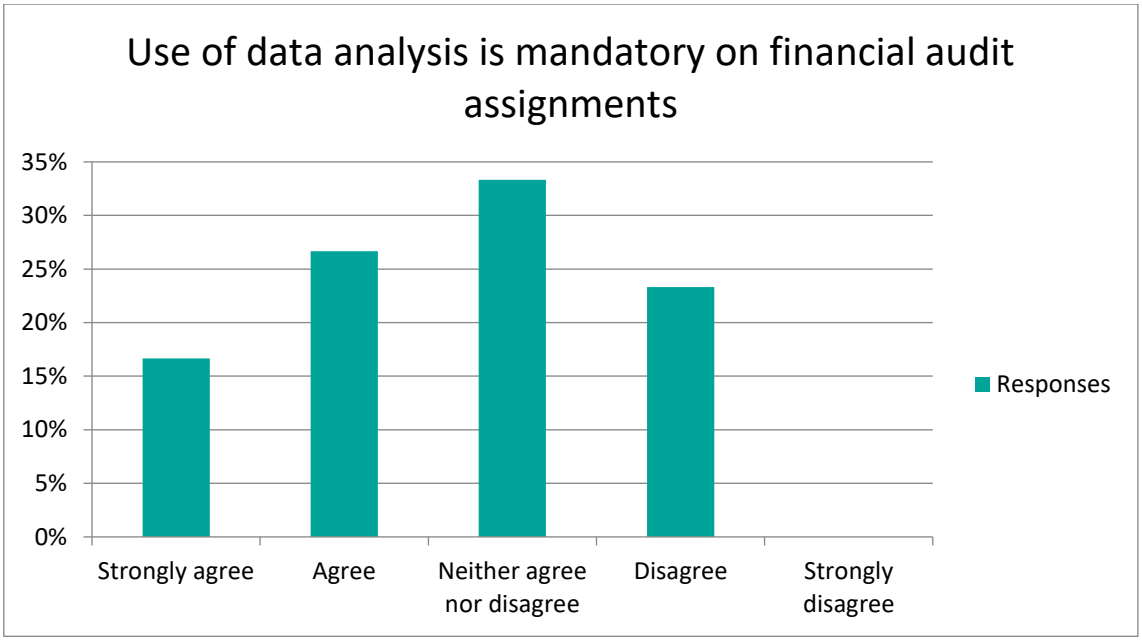

Figure 4-17 Use of data analysis in financial audit assignments

Less than half of the respondents considered data analytics mandatory in financial audit. Whether this was due to the SAI not having clear policies, or lack of knowledge, access to data or tools, is unclear. However, it gives an indication that many SAIs in the region are falling behind in reaping the benefits of the use of data in audit.

Initial inquiries on more complex data analysis conducted in SAIs revealed that this is something predominantly done by performance auditors. The study therefore sought to inquire about this data analysis capability. We see the same results here as for financial audit with only 40% agreeing that performance auditors had the necessary skills to apply advanced data analytics (3% strongly agreed and 37% agreed), while 30% neither agreed nor disagreed that this capacity existed in performance audit functions. 27% and 3% of respondents disagreed and strongly disagreed, respectively. See figure 4-18 below.

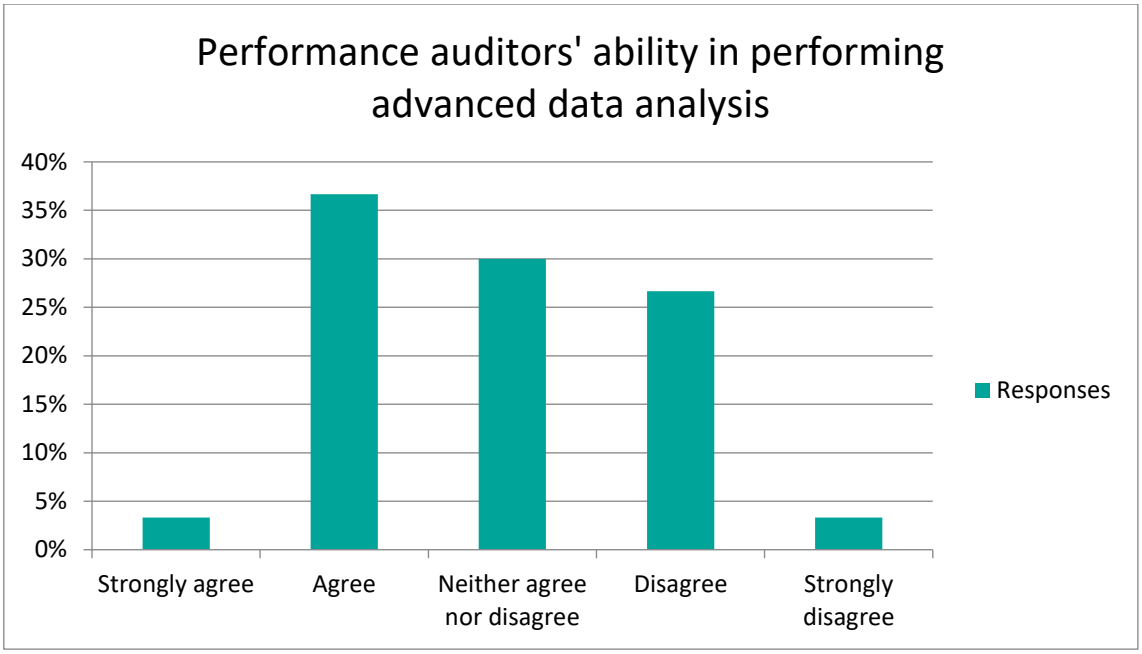

Figure 4-18 Performance auditors' ability in performing advanced data analysis

On the issue of whether the SAIs had or were able to get (big) data on which to apply advanced analytics, 20% of the respondents strongly agreed and 50% agreed. 10% neither agreed nor disagreed, while 13% and 7% of respondents disagreed and strongly disagreed, respectively. See figure 4-19 below. These results may be anecdotal because few or none of the SAIs have attempted, or are planning, to conduct big data analytics.

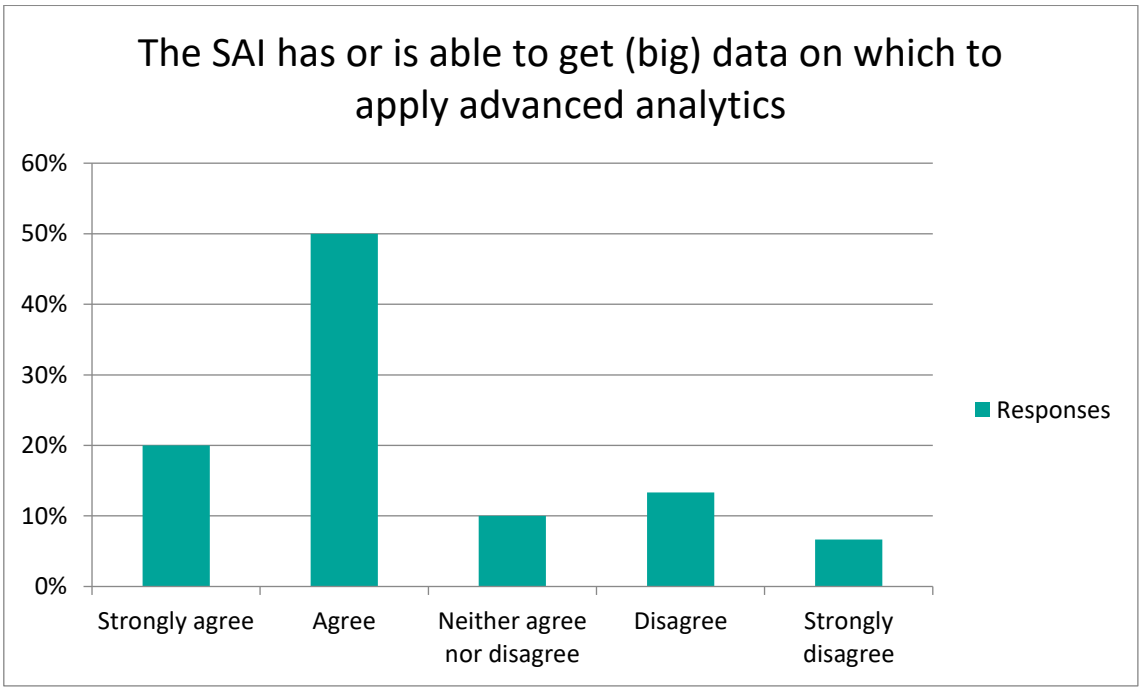

Figure 4-19 SAIs' access to big data

4.4.3 SAIs' capacity

SAIs should carry out a self-assessment of their internal capacity to deal with big data. However, 67% of the respondents indicated that their SAIs had not assessed requirements for the use of big data in public sector audit. Only 33% of the respondents claimed that their SAIs had conducted such an assessment. See figure 4-20 below.

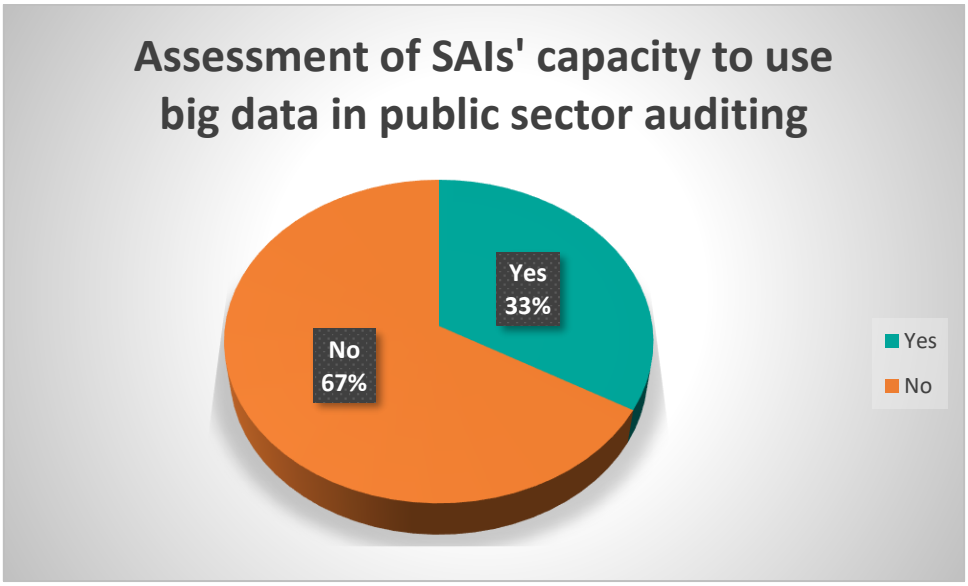

Figure 4-20 Assessment of SAIs' capacity to use big data in public sector auditing

4.4.3.1 People

With respect to capability, new skill sets may be required to support emerging analytics capabilities. SAIs need to review the skills of their employees and if they do not have these skills in-house already, they may need to recruit accordingly. However, data science skills are in demand and SAIs may not be able to attract and retain competent staff. The HR functions should be equipped to plan and strategise for future capabilities. Further, as the adoption of big data analytics is a huge investment for an organisation, the tone at the top is important to ensure understanding and acceptance at all levels.

On whether the SAIs had the PEOPLE in place to build capabilities that would make productive use of data that the SAI collected, 3% of the respondents strongly agreed and 53% agreed. 23% neither agreed nor disagreed regarding the availability of human resources. 17% and 3% of respondents disagreed and strongly disagreed, respectively. See figure 4-21 below.

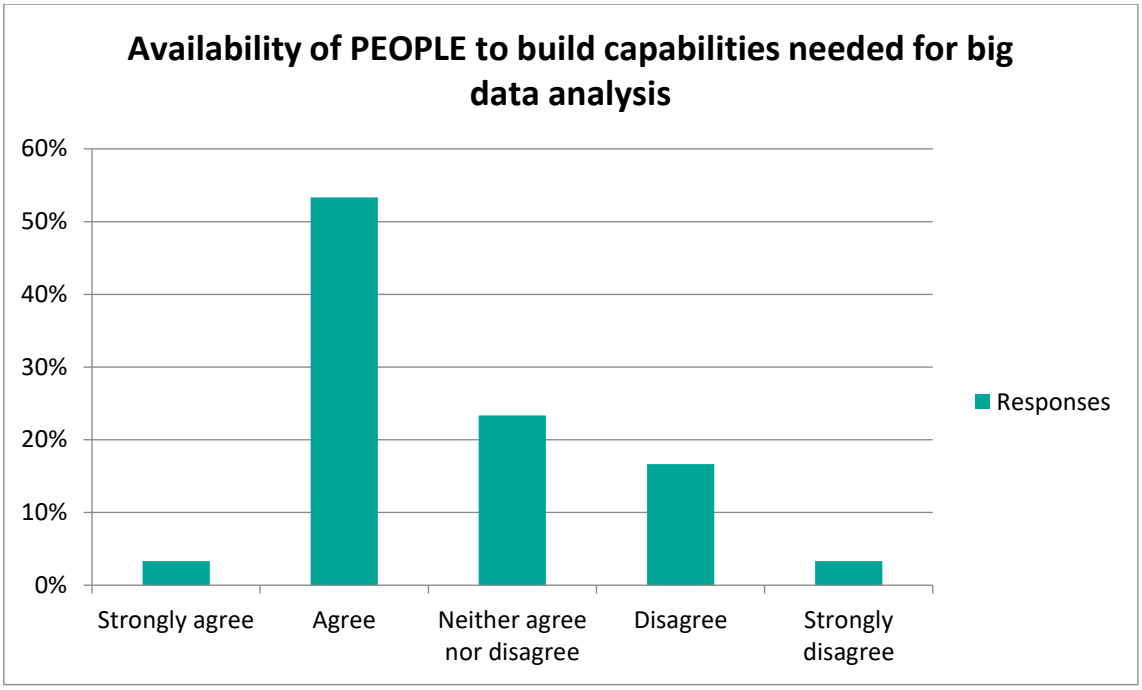

Figure 4-21 Availability of PEOPLE to build capabilities needed for big data analysis

On the availability of data scientists or specialised data analysis experts who conduct data analysis on large data sets, only 23% of respondents agreed. 77% of respondents indicated that their SAIs did not have experts. See figure 4-22 below.

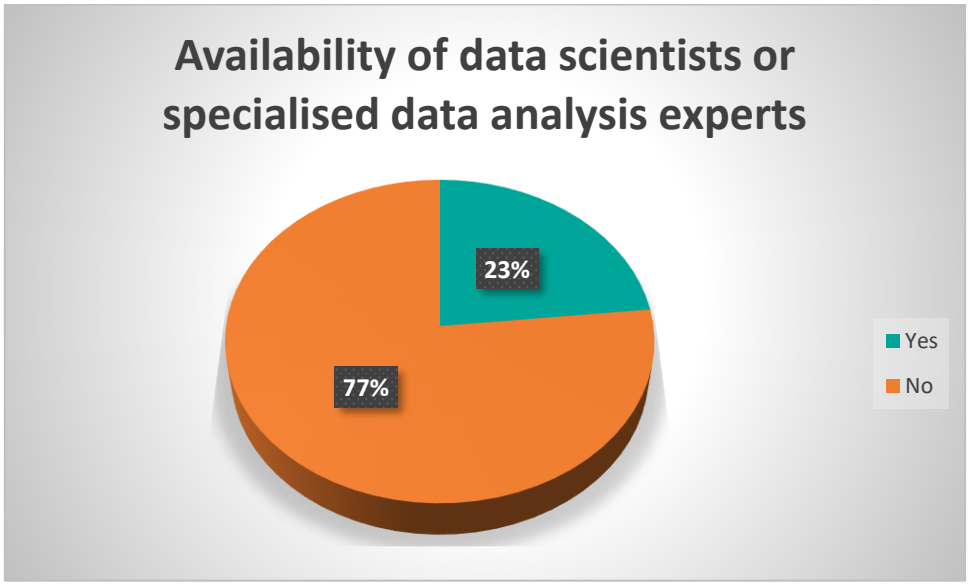

Figure 4-22 Availability of data scientists or specialised data analysis experts

Compounding the issue of lack of experts is the lack of dedicated teams for data analytics in SAIs. Only 7% of respondents said that the SAI had a dedicated data science or data analytics unit. See figure 4-23 below.

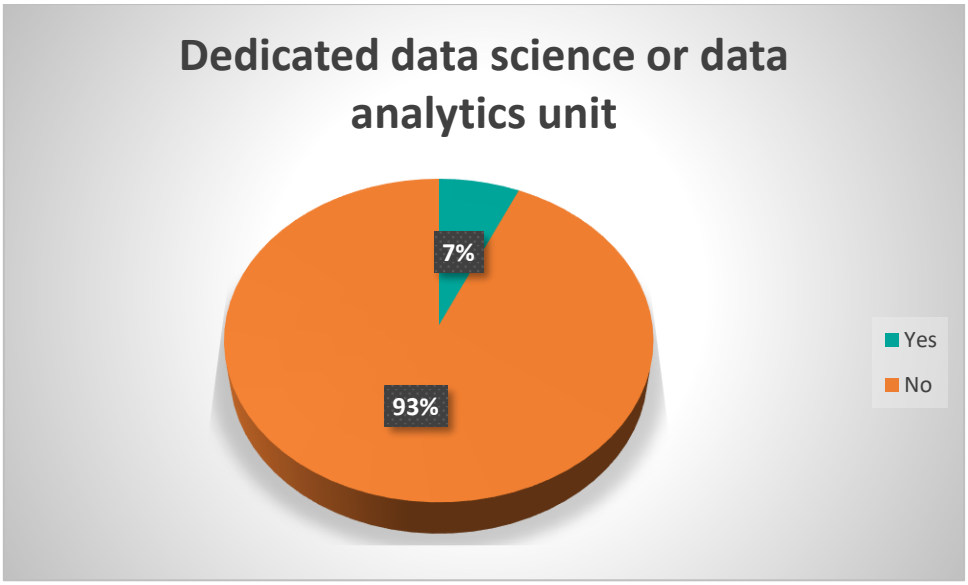

Figure 4-23 Dedicated data science or data analytics unit

The responses in figures 4-22 and 4-23 seem to contradict figure 4-21. The comments given by the respondents suggested that the SAIs have people who are available and willing to be trained to be champions for a data-driven SAI. These people need to be developed, provided with the right tools and supported by management.

4.4.3.2 Processes

From a process standpoint, the SAIs' access to government data is determined by the government's efforts to manage and share data within the public sector. SAIs need to consider that MDAs may not share data currently and may have a history or culture of resisting efforts to publish data openly, as well as a lack of comprehensive data governance. These barriers can impede open and collaborative exchange of important data elements and act as a barrier to adaptation in response to conclusions drawn from analysing the data. Lack of oversight and data governance may result in key stakeholders not knowing precisely where key data elements reside or how to access those data elements. Likewise, technology adopted without centralised oversight can complicate information sharing because it may represent a significant repository of critical information, and lack of central awareness of the information may limit the ability to include it in the scope of government data. Further, lack of data collection by governments may also pose a challenge.

Access to data also extends beyond the public sector's administrative data. Private interests, such as internet and mobile phone service providers, utilities, social media, financial and retail companies presently generate large volumes of data. Big data is a valuable commodity to these companies, either as their key product or by enabling them to gain a competitive advantage[43]. Gaining access to such data requires governments and SAIs to form partnerships or create legal instruments to compel companies to share data. Neither approach will be easy to implement.

[43] Kitchin, R. 2015, June. *What does Big Data mean for official statistics?* National University of Ireland, Maynooth

3% of the respondents strongly agreed and 63% agreed that the SAI had the PROCESSES in place to build capabilities that would make productive use of data that the SAI had collected. 27% neither agreed nor disagreed. Only 7% of respondents disagreed. See figure 4-24 below.

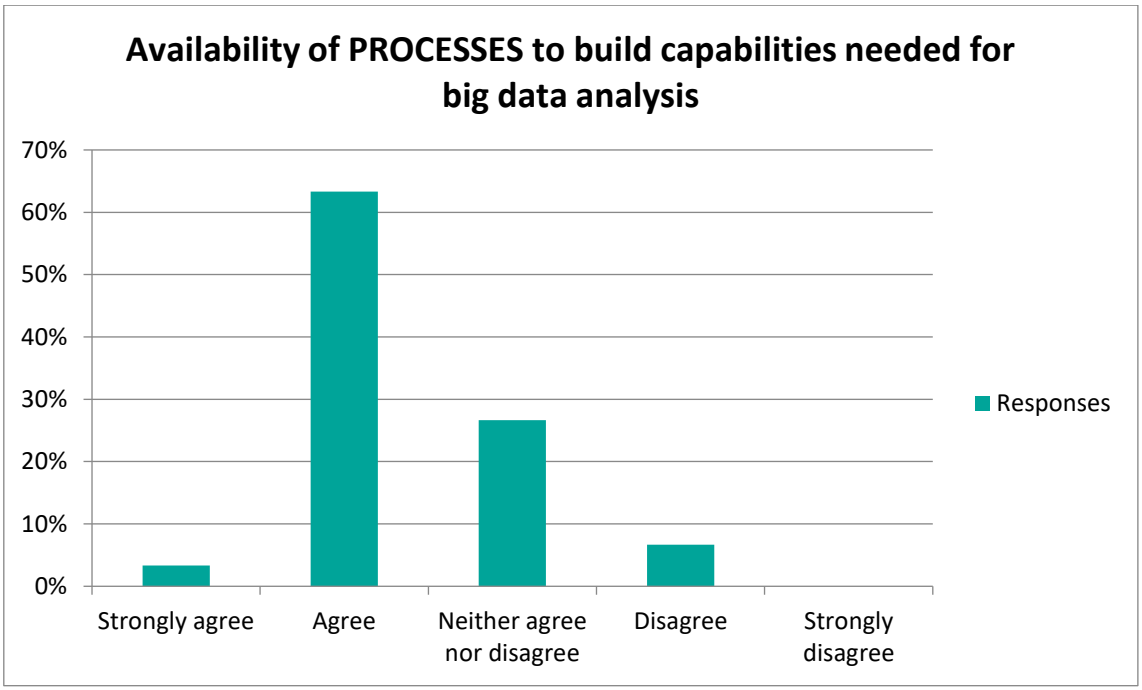

Figure 4-24 Availability of PROCESSES to build capabilities needed for big data analysis

Very few SAIs in the region have made advances in data analytics, and the processes around obtaining the data are often manual. AFROSAI-E has not seen any examples where a SAI has automated the process of transmitting data from clients to their office. Further, it is customary for each audit team to be responsible for obtaining the data sets they need. The few comments given with the responses suggested that the respondents may not have understood the question. The processes mentioned are standard scripts to analyse financial data and training programs to develop capabilities in data analytics.

4.4.3.3 Technology

The implementation of technology plays a role in determining organisational readiness. In many cases, new tools are required to support data analysis, and SAIs may need to evaluate capabilities for data storage and computation sufficiency. Moreover, sufficient data on which to operate needs to exist and be accessible to analysts. Data sources must be identified, which involves locating structured data (e.g., data organised in a relational database) and unstructured data (e.g., data stored ad hoc on a file system or in a loose collection). Identifying data sources can likewise involve data in a variety of different formats, including video, audio, images and text. Computational resources may need to be expanded to enable operation and analysis of these data types. SAIs must also assess operational and

technological risks, like privacy and data protection regulation, access to and integrity of data, before implementing big data analysis.

Because of these factors, big data analytics can amount to a significant investment that requires the involvement of many departments of the organisation, like management, HR, IT and audit departments. SAIs are required to think through supporting processes and identify potential problem areas before undertaking any investment.

The implementation of technology has previously been seen as a challenge in the AFROSAI-E region[44]. This is predominantly caused by budgetary constraints.

SAIs were asked about their technological capacity to utilise big data in audits. 3% and 26% of the respondents strongly agreed and agreed respectively, that the SAI had the necessary technology in place to utilise big data in audits. 33% neither agreed nor disagreed, whereas 33% and 3% of respondents disagreed and strongly disagreed. See figure 4-25 below.

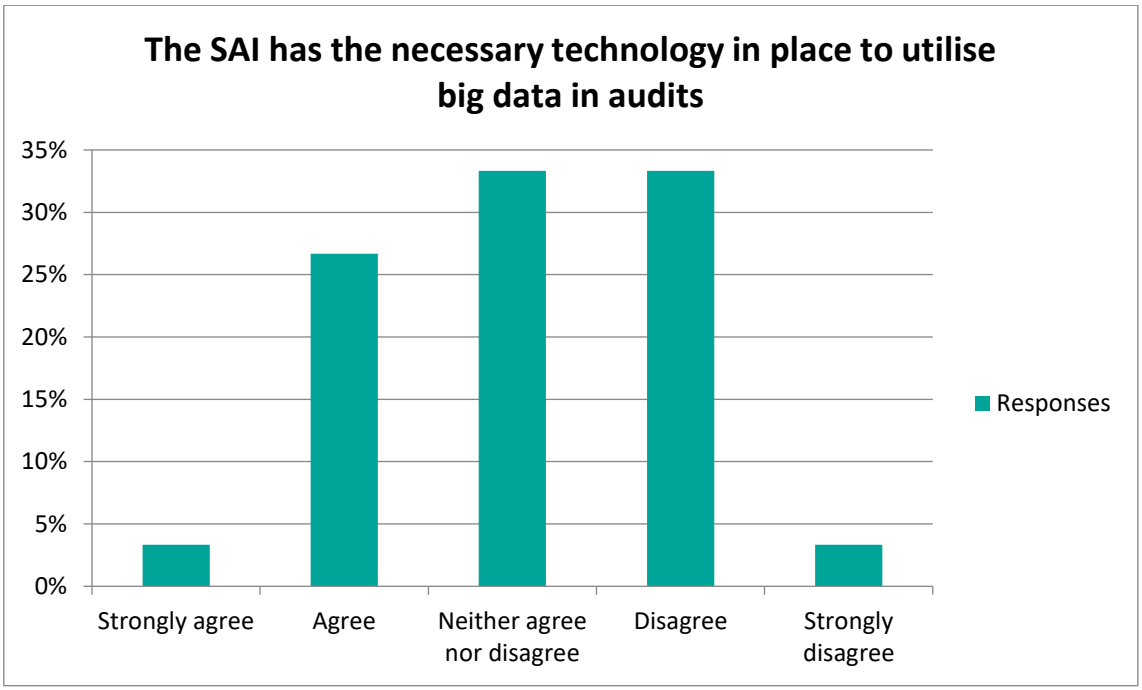

Figure 4-25 The SAI has the necessary technology in place to utilise big data in audits

Although almost 30% of the respondents (strongly) agreed that their SAI had adequate technology, their comments reveal that this may not be the case. Many SAIs may lack storage, bandwidth and analysis software. SAIs tend to utilise some sort of computer-aided tool. When asked if the SAI had other tools than MS Excel that they used for data analysis, 57% of the respondents agreed. The tools being used are Stata, ACL, IDEA, SPSS, Arbutus Analyser. See figure 4-26 below. The comments,

[44] AFROSAI-E. 2020. *2019 State of the Region ICBF Self Assessment Report.* Pretoria: AFROSAI-E.

however, revealed that many SAIs struggle with paying licence fees for their software, which will hamper advances in data analytics.

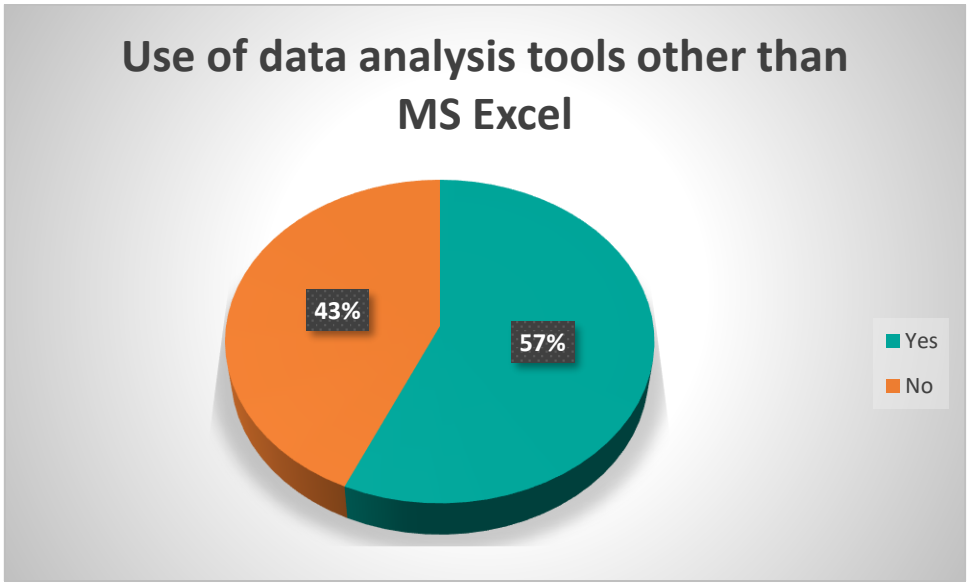

Figure 4-26 Use of data analysis tools other than MS Excel

To assess the severity of not having data analysis experts, SAIs were asked to indicate whether there were instances when certain audits could not be conducted well because of lack of data analysis skills. 67% of the respondents agreed that the SAI was **not** able to conduct certain audits either because of lack of data, or lack of skills in data analytics. Only 33% disagreed with this. See figure 4-27 below.

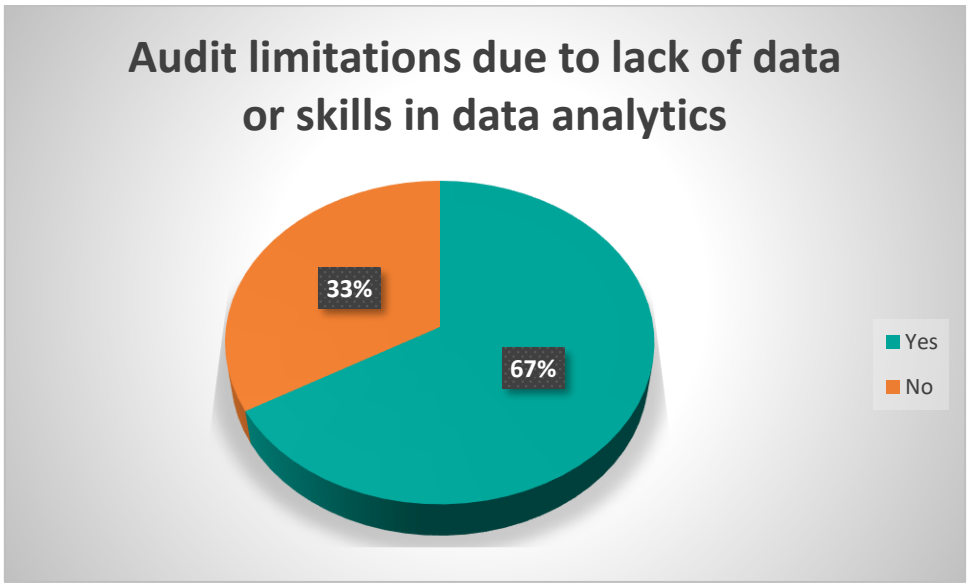

Figure 4-27 Audit limitations due to lack of data or skills in data analytics

4.5 Future considerations and guidance for the use of big data in public sector auditing

More than 67% of the respondents had suggestions on how big data can be used in public audit. This is promising, as it shows that the SAIs are seeing the potential uses of big data.

On whether there were significant opportunities to utilise big data analytics in public audit and government in general, 13% strongly agreed and 57% agreed. 23% neither agreed nor disagreed, whereas 3% either disagreed or strongly disagreed. See figure 4-28 below. Refer to appendix 6.3 for a full list of opportunities provided by respondents.

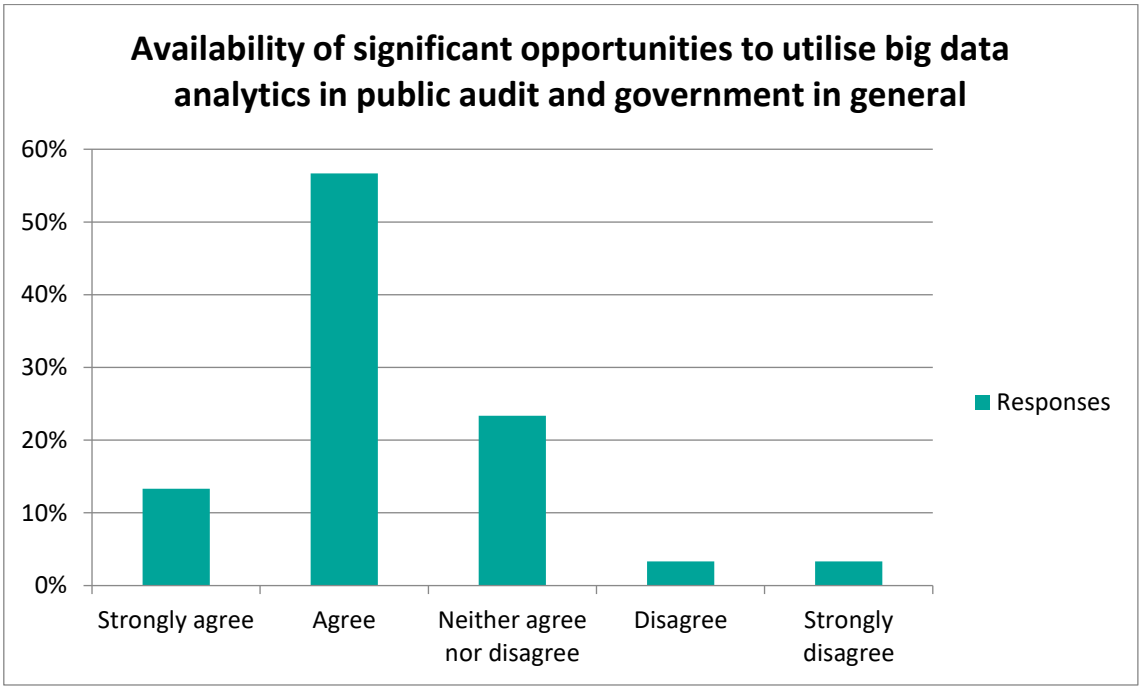

Figure 4-28 Availability of significant opportunities to utilise big data analytics in public audit and government in general

29% and 54% of the respondents strongly agreed and agreed respectively, that there were future uses of big data in public audit. 18% of respondents neither agreed nor disagreed. None of the respondents disagreed on the future uses of big data. See figure 4-29 below. These responses are consistent with the responses on the opportunities for use of big data discussed in section 4.3.3.

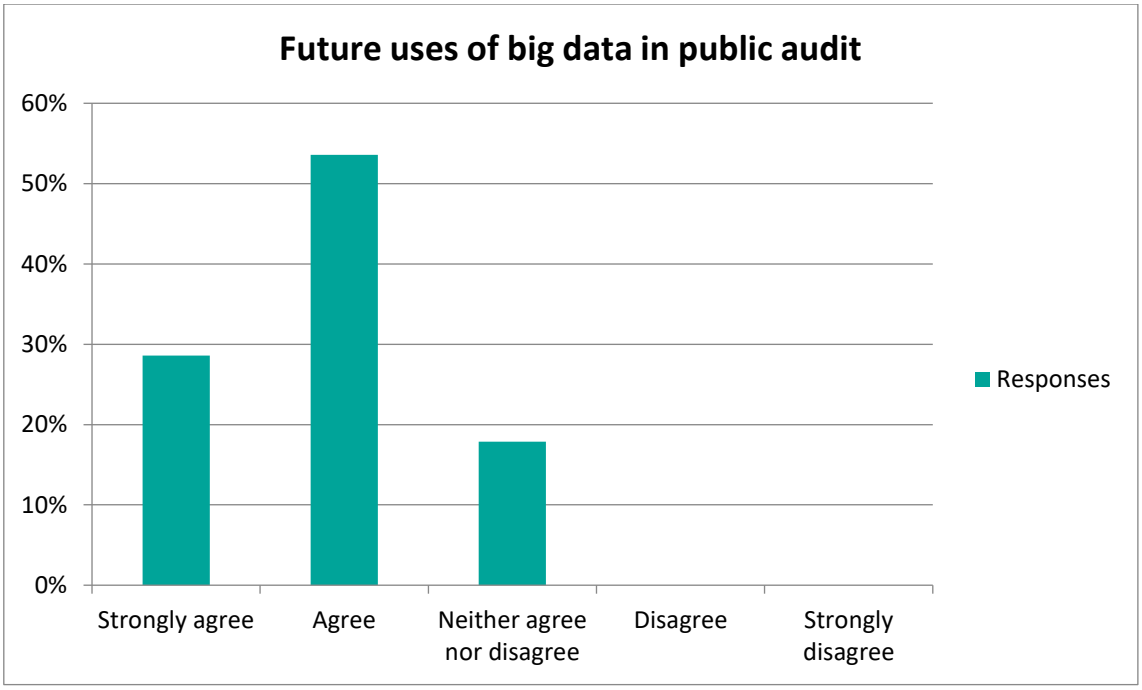

Figure 4-29 Future uses of big data in public audit

29% and 57% of the respondents strongly agreed and agreed, respectively, that there were future uses of big data in government. 14% of respondents neither agreed nor disagreed. None of the respondents disagreed on the future uses of big data in government. See figure 4-30 below. These responses are consistent with the responses on the opportunities for use of big data discussed in section 4.3.3.

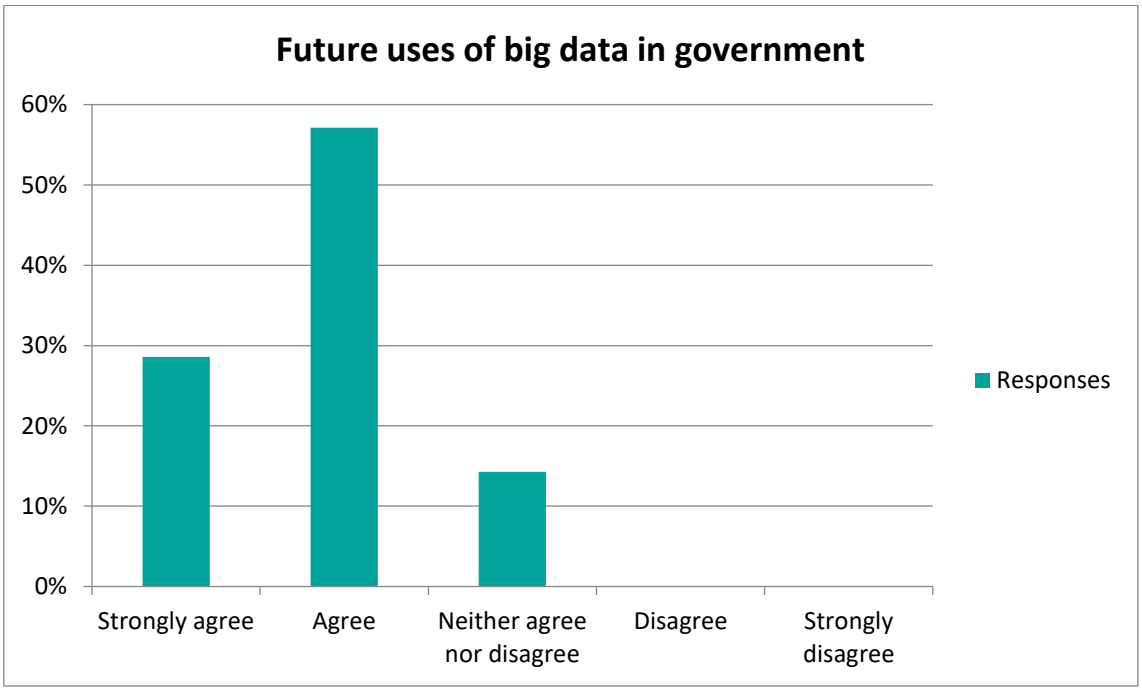

Figure 4-30 Future uses of big data in government

SAIs were asked whether they thought there was enough guidance currently for public auditors on the use of big data in public sector audit.

Only 7% and 11% of the respondents strongly agreed and agreed, respectively, that there was enough guidance on the use of big data analysis in public sector auditing. 29% of respondents neither agreed nor disagreed. 43% disagreed and 11% strongly disagreed that there was enough guidance. See figure 4-31 below.

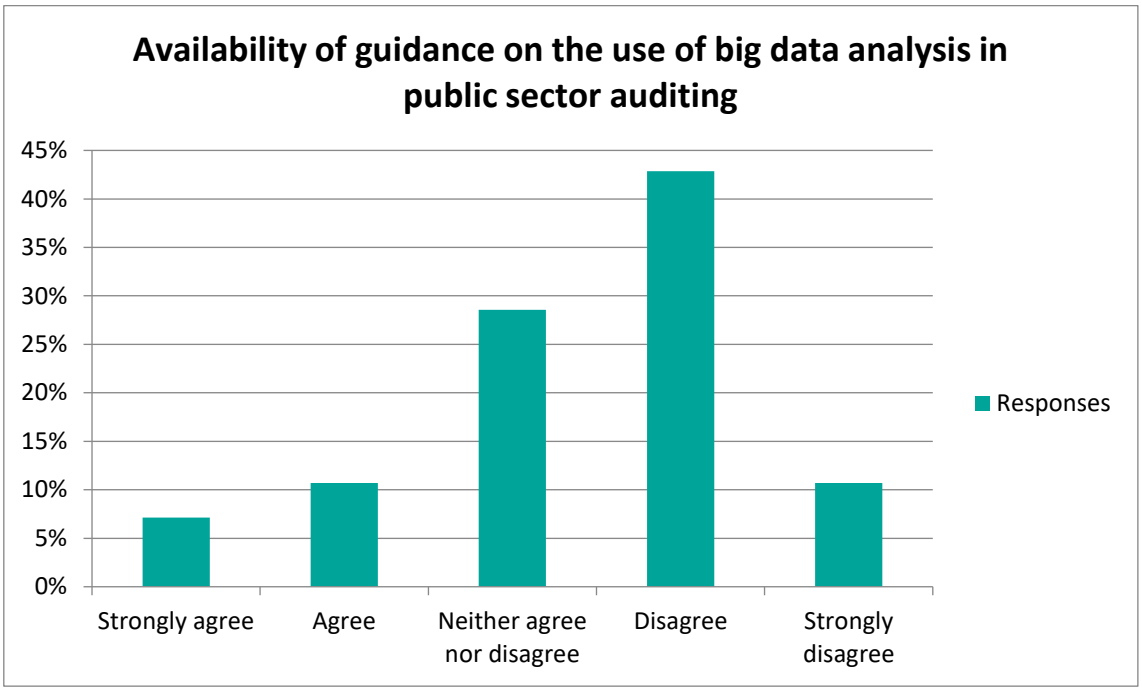

Figure 4-31 Availability of guidance on the use of big data analysis in public sector auditing

5 THEME STUDY SUMMARY AND RECOMMENDATIONS

5.1 Integration of big data in public sector auditing within AFROSAI

Levels of country computerisation and data generation

- ✓ Most of the respondents rated their country's level of digitalisation as satisfactory. However, other studies contradict this and show that African countries have a lower level of digitalisation. The respondents who stated that key processes and reporting in the public sector were not yet automated, further confirm this.
- ✓ Governments are increasingly realising the importance of ICT and implementing plans for utilising it. There is a need for SAIs and governments to look at country prioritisation of digitalisation projects. Adequacy of the legislative regime controlling or guiding ICT usage should also be reviewed.

Data governance

- ✓ Generally, government data is not easily available both within government and to the public. International studies show that African nations are lagging behind with regard to openness and access to government data. This can be challenging for SAIs that rely on data for their audits. However, this study indicates that the SAIs see opportunities within their own countries and those governments are making efforts to become more data driven.
- ✓ SAIs should put data on their agenda and in their strategies. It is important for SAIs to identify stakeholders and repositories of data, and form partnerships that will enable access to this data. Further, SAIs should promote openness of data and nudge governments to ensure that public data is available in a machine-readable and reusable format, easily accessible to the public online, and that datasets are up to date and kept regularly updated.

Big data application in government

- ✓ The understanding of big data is greater in the SAIs than in governments in general. However, overall the study reveals that there is not a clear distinction between big data and mere large volumes of data and that the term is applied loosely. We do not see this as an obstacle because there is a need for SAIs and governments to become more data driven; whether it be big data or not, is of less importance.
- ✓ A majority of the respondents agree that big data presents an opportunity to government. Although the utilisation of (big) data for policy-making and delivery of services is still in its initial stages, the SAIs report that many governments recognise the value and benefits associated with the utilisation of data.

SAIs' use of data and big data in audit

- ✓ Our study shows that data analytics is used in most SAIs in the region. However, we recognise the need to strengthen the capabilities for auditors in all three of the audit types, and the need to build capacity in more advanced data analytics.
- ✓ Analysis of financial data in the audit of financial statements is the predominant form of data analysis in the region. The respondents report that they have access to data for this purpose.
- ✓ Although the SAIs should form partnerships beyond the public sector with regard to obtaining relevant data, the main data sources for SAIs will be their auditees. Whilst over half of the SAIs have identified stakeholders, it seems that they have not taken the appropriate steps to form partnerships to ensure access to relevant data beyond financial data.
- ✓ The response on the ability of the SAIs to get big data on which to apply analytics is consistent with their response on how government data is accessible to the public auditors. However, data analytics is still in the initial stages in most SAIs in the region, hence the SAIs may face challenges that they did not foresee when expanding the extent and moving to more advanced data analytics.
- ✓ Despite the respondents' optimistic view of the SAIs' readiness for big data analytics, the SAIs in the region do not have the people, the processes nor the technology to build capacities in big data analysis. Very few SAIs have assessed their internal capacities, which implies that there is not proper buy-in from the top. The study shows that there is a lack of expertise in the data field. Several SAIs in INTOSAI are employing data scientists and specialists who are leading the SAIs' data analytics strategies. It seems in AFROSAI this need is not yet recognised as important, which could explain why these SAIs are lagging behind other INTOSAI members with regard to data analytics. Processes to identify data sources and form partnerships are immature. Further, the analysis tools that the SAIs currently possess are not built to handle the volumes and formats that big data require. One of the notable challenges big data brings is that traditional systems and tools are not adequate to store, process and analyse the data. Hence, the need for varied tools and strategies is greater.

5.2 How SAIs can overcome big data challenges

The SAIs should have a clear strategy for data analysis in all three types of audit, which will ensure buy-in throughout the organisation. The strategy should include the following aspects:

- ✓ Direction and impact: The utilisation of data in audit has to have a purpose. The SAI should assess how data analytics can be integrated in the three types of audit and what its desired impact should be. The assessment should also identify risks that have to be addressed in order for the SAI to achieve its goals.
- ✓ People: To become more data driven, the SAI has to assign clear roles and responsibilities. The HR department has to assess whether the needed skills can be found in-house or whether recruitment is needed. Training programs for staff may have to be designed and implemented, for both auditors and expert data analysts.

- ✓ Processes: The SAI should identify stakeholders and form partnerships with both government institutions and external institutions that maintain relevant data repositories to ensure continuous access to all relevant data. Where there is a lack of data, or government fails to update or keep data in machine-readable and re-usable formats, the SAI needs a strategy to find alternative data sources. The SAIs should also include any shortcomings in government data handling in their reports and recommendations. Processes to transfer the data from the auditee/data partner to the office and distribute it to relevant audit teams will have to be designed and implemented.
- ✓ Technology: Through its assessment of its capabilities, the SAI may find that its infrastructure with regard to network, storage, processing power and analysis tools is not adequate. Investments in technology may be a significant cost, and a plan for scaling towards the desired level should be included in the SAI's strategy.

6 APPENDICES

6.1 Appendix 1 – Survey questionnaire

Integrating Big Data in Public Sector Auditi

Integrating Big Data in Public Sector Auditi

6.2 Appendix 2 – Opportunities to use big data analytics in government services

1. Data analysis in auditing.
2. Most of key data like births, deaths, citizen registrations, livestock registrations are easily maintained in the systems and easily accessed when auditing. Generally, most government services are maintained systematically and data banks maintained and backed-up for risk management.
3. The review of utilities agencies and some government corporations has opportunities for big data analytics.
4. Regulatory compliance through the management of natural resources such as water and forests.
5. Law enforcement.
6. Social services (social security and childcare).
7. Tax fraud.
8. Health services.
9. Information security, power supply, water supply, travel and tourism, inter-agency databases like security to curb crime, public health, national disasters like cyclone Idai, Covid-19.
10. The equitable distribution of services among citizens.
11. Most government agencies have stand-alone systems with very few (if any) computerised systems.
12. To protect the citizens.
13. Look for duplicate payments among different vendors on the IFMS. Seeing if there are ghost employees on the payroll which were declared dead on another system.
14. Health information, national population census, socio-economical information, agricultural information, etc.
15. Government has a lot of data that could be used by big data analytics.

6.3 Appendix 3 – Opportunities to use big data analytics in public sector audit

1. Data analysis, finding patterns of things, for example information collected, government ministries and agencies of any size can make data-backed decisions i.e. cyber security improvements, lower traffic congestion and improve the environment etc.
2. It is important to equip the users and reviewers with the required skills.
3. Before embarking on data analysis, it is important to train or equip users with the tools of data analytics.
4. The use of big data analytics in public sector audits will create new opportunities for auditors and make them capable of combining their routine audit skills with analytics to analyse industry data. In this regard, I suggest the following: (i) a dedicated data science or data analytics unit should be set up by SAI's. (ii) big data tools such as: Hadoop which helps in storing and processing large data, Spark - helps in-memory calculation, Storm - helps in faster processing of unbounded data, Apache Cassandra - provides high availability and scalability of a database, MongoDB provides cross-platform, should be acquired and put to use by SAI's.
5. MongoDB provides a cross-platform and should be acquired and put to use by SAIs.
6. To inform policy and government decisions.
7. Use procurement data analytics to identify unusual or suspicious transactions for audit or further investigation. Use of analytics to monitor financial transactions or test the effectiveness of internal controls and compliance procedures.
8. To conduct performance audits on the extent of government service to citizens, e.g. power and water supply, road infrastructure (from tolling data and accidents statistics). Big data analytics can inform the emergency/disaster response preparedness of a nation.
9. Statistical tools.
10. There should be integrated systems among government agencies that enable sharing of data
11. An indication of the rate of financial corruption.
12. Confirming the accuracy of one data source with another one.
13. Used to establish non-compliance and control weaknesses across government entities, especially regarding payroll-related risks.
14. Identifying double payments to public servants for cross-cutting activities, easy assessment of inter-governmental settlements and consolidated financial statements.
15. Review of budgetary controls across the country.
16. Big data in the public sector can be used for planning, monitoring and evaluation.
17. The government uses the IFMIS to process financial transactions in over 900 ministries and agencies across the country. The SAI needs to know how to analyse all this information for review purposes. The government also uses an integrated payroll system to process all salaries and allowances for all personnel working for the government.
18. The SAI needs clarity on what big data is and what opportunities it presents.

6.4 Appendix 4 – Comments

Comments - Integrating Big Data ir