Data deficits and democratic processes: The under-explored role of data in African elections

Guy Berger and Liz Orembo

RESEARCH ICT AFRICA

Workshop 17
17 Dock Road
V&A Waterfront
Cape Town, 8001
Cape Town, South Africa
Phone: +27 21 447 6332
Fax: +27 21 447 9529
www.researchictafrica.net

Acknowledgements

Authors: Guy Berger & Liz Orembo

Principal investigator: Alison Gillwald

Project manager: Naila Govan-Vassen

Research editor: Alan Finlay

Proofreader: Drew Haller

This work was made possible with a grant from the International Development Research Centre (IDRC), Ottawa, Canada. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

Citation: Berger, G., & Orembo, L. (2025) *Data Deficits and Democratic Processes: The Under-Explored Role of Data in African Elections.* Cape Town: Research ICT Africa

Copyright

First published by Research ICT Africa in March 2025.

Publisher

Research ICT Africa

Workshop 17

17 Dock Road

Cape Town

South Africa

8001

+27 21 447 6332

info@researchictafrica.net

www.researchictafrica.net

Executive summary

This report examines the perceptions of relevant stakeholders in regard to the elections data ecosystem with insights from South Africa, Kenya, Ghana and Zimbabwe. It is based on the premise that election integrity relies on an environment of information integrity, and equally – but less visibly—on data integrity and availability. In this context, ensuring that information and data are accessible, reliable, and treated is critical, given that it directly influences the transparency and fairness of elections. The research problem addressed is a knowledge gap about how the primary electoral stakeholders see data issues across the diverse dimensions of democratic processes, and the relevance of this to wider policy issues around access and use of public and private data in African countries.

The study methodology combines a literature review and interviews with key African election stakeholders, supported by case studies, the monitoring of webinars, and convening an in-person discussion. Sources include elections management bodies (EMBs), election analysts, political parties, civil society organisations, media stakeholders, and tech companies, thereby providing diverse perspectives on the elections data ecosystem.

The research finds some recognition of the importance of data availability and integrity in the election process. However, the centrality of these issues to credible democratic outcomes is not fully understood and acknowledged among most stakeholder groups. Actual engagement with the range of data challenges and opportunities is limited, and substantial silos exist between the different actors. Further, despite calls to social media platforms for meaningful access to data to monitor African elections online, this prospect remains largely unrealised.

The report offers suggestions for moving forward, particularly at the policy level, that could help to improve the use of data to ensure election integrity on the continent. It acknowledges the significance of the African Commission for Human and Peoples' Rights' November 2024 resolution as a major step towards improving data access and use for election integrity. The main recommendations include increasing stakeholder awareness through multi-sector roundtables to encourage collaboration and address fragmentation; strengthening governance structures within EMBs to balance privacy and data availability; developing and implementing data access policies; and advocating for digital platforms to provide granular data access for researchers, journalists, and fact-checkers for improved election monitoring.

Key findings

1) Limited awareness and understanding of the value of data as information for public good
    - The value of data for elections, and the advantages of accessing, analysing and sharing it, is not holistically perceived or assessed by stakeholder groups; and
    - Legal regimes for data protection across African countries are more widespread than for information and data access, but balancing privacy and access is not prominent on stakeholder agendas. Data protection does not work in many parts of Africa due to lack of

capacity and other political factors, while right to information regimes are limited in regard to data access.

2) Statutory stakeholder and data access

- On the part of state institutions, including EMBs, there are attitudes of secrecy and a lack of understanding of the obligations to make data available. Some fear that sharing data can expose officials to legal action. Some authorities do not trust that openness will not be abused, which resonates with perspectives on how data is treated in authoritarian contexts.
- Turnover in government personnel associated with political regime changes or election cycles may frustrate policy implementation with respect to sharing data. Political appointees may want to implement data policies in the interests of their parties, rather than in the public interest, potentially distorting intended policy outcomes; and
- There is a lack of shared understanding of data issues and the role of data during elections among official bodies, including EMBs and other regulators.. This also applies to the classification of data in that certain data that these entities would not want to release is not consistently categorised as such.

3) Data access challenges

- Civil society and academia have to buy access to market research, opinion polls and platform data, and they often cannot afford the cost;
- Political parties need voter registration information to inform their campaign strategies, but apart from opinion poll results, they do not do much to source and use other data; and
- Telecommunications operators, according to some interviewees, breach data protection laws and too easily share data without legal compulsion which supports authoritarian actions by states during elections. Sometimes data accessed from the private sector, even within legal frameworks, can be abused and used for illegal and authoritarian purposes.

4) Institutional capacity

- There is also a lack of technical skills and capacity among EMBs, non-governmental organisations (NGOs), and the academic sector in accessing, processing, and sharing data relevant to elections. Furthermore, availing data that is format-friendly sometimes has cost implications and also requires dedicated hardware and software. Where data sharing is not legally mandated, institutions and other actors are not motivated to share their data publicly.

5) Political influence/interference issues

- There is a view among some interviewees that some data may be culturally sensitive and should be protected and only shared with the appropriate safeguards to preserve indigenous communities and protect their rights. In some countries, African voting patterns often reflect a pattern of ethnic affiliations with related data being contentious;
- Data plays into political power in several ways. It can negatively impact journalists, media houses as institutions, social justice activists, opposition candidates, and female candidates, amongst others, making them vulnerable to authoritarian regimes and online trolling from party supporters. It can also adversely impact the prospects for fair competition among political rivals; and
- Noting the gaps, different initiatives are emerging for data sharing, awareness-raising, and capacity building. For example, the African Open Data and Research Initiative, an NGO in Ghana, raises awareness of open data and provides technical support and advice to

communities and organisations in need of open data assistance. The African Alliance for Access to Data also does on-going awareness-raising and advocacy.

Table of contents

Introduction to the report

1. Background

This report and its underlying research reveal stakeholder perceptions about the significance of access to data in elections in Kenya, Zimbabwe, South Africa and Ghana. While issues of developing and dissecting data, who holds it, and who has access to it differ in each country, some commonalities exist and are relevant for other African elections. The findings highlight how data is at the heart of democratic processes. They further show contradictory sides of the equation. On the one hand, recognition and management of data can enable meaningful information and knowledge that promote and protect electoral integrity. On the other hand, abuse or mishandling of data can cause major damage to the credibility, accountability and stability of outcomes and distort the will of an informed electorate.

Appendix A provides an analysis of the data issues for elections, using a supply-and-demand perspective. This also reflects on African experiences in comparison to other parts of the world where there are greater levels of datafication and use. Sub-section 2 notes the relevance of African instruments to the topic. However, food for thought comes from one of the interviewees expressing reservations about current uses of data in African polls and whose perspectives are summarised in the box below.

A cautionary perspective

According to one of the interviewees, in semi-authoritarian contexts, data is not especially influential in terms of elections. A lack of transparency and data serves those in power, according to this respondent. Further, underlying politics is often opaque and offline; what transpires online is in a different realm. For this interviewee, the limited data that is available cannot be verified, and is used to manipulate processes. Opinion polls are often inaccurate, the person believes, but are nevertheless exploited by whichever actor sees an advantage. Meanwhile, the public does not attribute much importance to such data-generated findings.

Noting this view, this report nevertheless recommends stakeholders to consider increasing their preparedness for the likelihood that, following global trends, African elections are likely to become more data-driven over time.

2. Policy overview

At both national and regional levels, Africa has a number of policies and laws that can be used to facilitate access to data within the elections data ecosystem. Out of 54 African countries, 29 have enacted access to information laws to enhance transparency in governance. Although these mainly lack explicit reference to access to data, these have potential significance for the matter. A total of

36 countries have data protection laws. Attention to both access and privacy issues is vital to ensure data as a public good.

At a regional level, the African Union (AU) has made significant efforts to establish frameworks for data governance which can be used in the context of elections and democratic processes. Recognising data as a strategic asset that can drive sustainable development and facilitate evidence-based decision-making, the AU Data Policy Framework (AUDPF) adopted in 2022 aims to enhance data sharing within and across Member States while promoting human rights, freedoms, fairness and inclusion. In conjunction with the AUDPF, the Malabo Convention also has significance for enhancing data governance on the continent. It was recently ratified after a prolonged process, establishing legal standards for cybersecurity and data protection, reinforcing some of the critical conditions necessary for data sharing.

In October 2024, UNESCO's conference on the International Day for Universal Access to Information, meeting in Accra, Ghana, agreed to the statement "Harnessing the Power of Data: A Commitment to Strengthening Access to Information in the Digital Age", which in turn followed pan-African and international consultations by a network called the African Alliance for Access to Data. This momentum continued in November when the African Commission of Human and Peoples' Rights (ACHPR) adopted the resolution *Promoting and Harnessing Data as a Tool for Advancing Human Rights and Development in the Digital Age*.[1] The resolution notes the foundational milestones of the Model Law on Access to Information for Africa, the Guidelines on Access to Information and Elections in Africa, and the Declaration of Principles on Freedom of Expression and Access to Information in Africa. The resolution states: "Data held by public institutions and bodies receiving public funds, as well as that held by private actors where there is an overriding public interest in access, should be made publicly available by default, in alignment with the principle of maximum disclosure, except where justified by regional and international human rights standards". The resolution additionally recognises the critical role of cybersecurity and equal access to data, as well as the role of private sector companies in harnessing opportunities presented by data. It further mandated the Special Rapporteur on Freedom of Expression and Access to Information in Africa "to consult broadly around the continent to examine and develop appropriate normative standards to guide data collection, deployment and access issues concerning data". This provides a relevant opportunity for electoral stakeholders to advance understandings about the issue of access to and use of data in this sphere.

The challenge is substantial. Digitalisation and access to data in Africa are very uneven, owing to factors that include uneven distribution of infrastructure,[2] different types of democracies, external ownership of relevant data holdings, insufficient laws and underlying economic and cultural

[1] ACHPR, (2024), Resolution on Promoting and Harnessing Data Access as a Tool for Advancing Human Rights and Sustainable Development in the Digital Age. ACHPR/Res.620(LXXXI) 2024, https://achpr.au.int/en/adopted-resolutions/620-data-access-tool-advancing-human-rights-and-sustainable-development, Accessed on 15th November 2024

[2] Research ICT Africa (2017), A Demand Side View of Mobile Internet from 10 African Countries, https://researchictafrica.net/research/after-access-2018-a-demand-side-view-of-mobile-internet-from-10-african-countries, Accessed on 15th November 2024

barriers. This unevenness has also fed into challenges of disinformation especially during democratic processes.

Cybersecurity remains a key component for trust between stakeholders in data ecosystems, yet Africa's cybersecurity maturity is generally considered low compared to other regions. According to the International Telecommunication Union's 2024 Global Cybersecurity Index, only 55% of African countries have national Computer Incident Response Teams (CIRTS) to monitor and respond to cybersecurity incidents, and only 25% have CIRTS for sectoral-specific cybersecurity responses.[3]

The AU policies on cybersecurity, data flows and digital trade have sought to harmonise policies across African countries to bridge inequalities between countries and promote trade within Africa. Significant policy efforts at the regional level on investments to support African countries to benefit equally from digitalisation are lacking and could be among the causes of the fragmented policies in the region which also have an impact on elections issues.

While these policy frameworks have the potential to promote open and rights-respecting data sharing for election integrity, their uneven and non-implementation leads to national policy uncertainties and a lack of trust required to unlock data and end data silos. The challenges at national level – as borne out by the interviews for this report which offer granular perspectives of stakeholders at the coalface of elections – include: a lack of political awareness and will, dysfunctional public services, a culture of official secrecy, authority over tech actors outside the jurisdiction, restrictive provisions in laws, and poor infrastructure.[4] A number of these challenges are deeply rooted in colonial legacies, as highlighted in several stakeholder interviews.

3. Outline of data types and stakeholder interests

Data sources relevant to elections are diverse and serve different purposes. They include:

- Statistics on official electoral registration, voting stations and voting results, as well as procurement records for Electoral Management Bodies (EMBs);
- Voter opinion surveys;
- Media, marketing, advertising and campaign spending data;
- Data from digital platforms that are relevant to elections; and
- Data from telecommunication operators that can give insights related to elections, such as network coverage and accessibility.

[3] ITU (2024), Global Cybersecurity Index, https://www.itu.int/en/ITU-D/Cybersecurity/Documents/GCIv5/2401416_1b_Global-Cybersecurity-Index-E.pdf

[4] Africa Freedom of Information Centre, (2024) Why Access to Information Matters for Africa: Exploring the Challenges and Importance of ATI Implementation, https://www.africafoicentre.org/overview-of-the-current-state-of-access-to-information-in-africa/, Accessed on 15th November, 2024

Most election stakeholders have an objective interest in data sets that pertain to an election. However, there are also varying primary roles and stakes as follows:

1. EMBs and election observers have the mandate to collect and disseminate official data on electoral issues, and to be transparent in their operations and contracting. There is also interest in network coverage for cases where election results transmission depends on the availability of connectivity;
2. Political parties and analysts tend to give special attention to opinion polls, attitudinal data, and voter registration, as well as to data about media coverage, advertising and campaign spending, including of rival contenders;
3. Journalists, fact-checkers and election monitors have a particular interest in data on digital platforms, such as sentiment trends and influence operations, as well as patterns showing co-ordinated inauthentic behaviours including disinformation campaigns and similar attempts to discredit elections;
4. Digital platforms use data points relevant to their content moderation and their advertising businesses during election periods;
5. Voters' interests are mainly in information (rather than data), but much information is based on underlying data, including about the history and performance of parties and candidates in the polls;
6. Telecommunications companies hold data that can give better insights into elections-related data including network connectivity and use, website traffic, and internet restrictions (throttling and shutdowns); and
7. Public relations and marketing companies, including paid influencers, are deeply invested in assembling and analysing data for their clients.

In this study, we conducted interviews with key stakeholders involved in various aspects of elections. Our participants included representatives from the EMBs, think tanks, civil society organisations, media outlets, Pan-African organisations, and election observers. Some of these stakeholders have interconnected roles. The purpose of our inquiry was to understand data gaps as well as who utilises what data, who is responsible for generating it, who owns and controls it, what data is in demand, and the methods of accessing, processing and sharing data (which may be beneficial to other stakeholders). The methodology and research questions used can be seen in Appendix B.

2. Findings

2.1 Data issues and EMBs

EMBs are collectors and custodians of data that is mission-critical for electoral integrity. They depend on data gathered indirectly, as well as directly. An example of the former is data such as births/deaths and addresses to manage voter registration and to prepare voting infrastructure. EMBs depend a lot on getting data from state bodies, such as statistics agencies, public pensioner databases, youth and education ministries, and death registries. Where elections have technology

components and require network coverage at the polling centre, for example as with the real-time transmission of election results in Kenya, the EMBs need data from telecommunications operators, communications regulators and statistical commissions about relevant demographics. EMBs are further interested in data relevant to managing their reputations, given that the credibility of a poll depends on how they are perceived.

Regarding direct data, EMBs collect several data sets themselves. For example, during registration drives they may gather information about voters with special requirements such as the elderly or people living with disabilities, and/or they may collect biometrics. Some collect voter identification (ID) numbers and cellphone numbers, and share the latter with cellphone partners to send out public service messages. There will be data about procurement for the many external services that EMBs require and store, and there are sets they compile about the accreditation of media and observer missions. Assembling data on political parties is also important for EMBs where their role includes regulating political party funding.

Many data inputs for EMB use need to be cleaned, and most have to be stored and secured so there is a historical trail. However, there can be issues about the validity and reliability of EMB data. Interviewees noted that:

- Data pertaining to the delimitation of voting districts held by EMBs is not always transparent, and underlying census data is not always reliable. Voter registration data is not always up to date regarding the current addresses or living statuses of voters. There is an ongoing need to clean EMB data and to reinforce its security and confidentiality;
- Voter turnout figures are typically presented in relation to the number of registered voters, but detailed metrics for comparing turnout to the total number of eligible voters can be hard to find on an EMB's website. Data on voters' literacy levels is sometimes missing, although relevant to electoral education initiatives; and
- Political contexts where elections management officers are appointed by the executive arm of the state, and which is controlled by political forces with direct interests in the elections, raise trust issues in the credibility and handling of data.

Interviewees also remarked on the availability of EMB data as an issue:

- What data should be made available to the public entails policy positions and decisions, also with a view to adhering to privacy legislation. One EMB lets voters check their registration data, for example, and while it collects voter ID numbers, it replaces numerals in the data sets with asterisks so as to protect privacy;
- In one country, data on voter registrations was leaked to the ruling party, including the mobile phone numbers of voters, who then received campaign calls. The data protection authority was called on to investigate, but said it could not trace who was responsible. In the same country, the official release of the voter registration data (without personal information) was too close to the elections to be of practical campaigning value. It was also given exclusively to political parties and in .pdf rather than machine-readable format;
- In another country, an EMB employee released the candidate lists of two parties which contained their ID numbers. The case was taken up by the Information Regulator who asked whether the incident had been reported on the EMB's website, if the leak had been

communicated to the parties concerned, and what mitigation measures had been implemented. In one instance, the EMB's website was not always online nor did it offer granular detail such as district voting patterns; and
- There are unclear policies about what candidate data (e.g. educational qualifications) should be legitimately in the public domain.

On technology and capacity:
- At least one EMB contracts out some functions relevant to data, notably the monitoring of print, broadcast and digital platforms. This EMB receives up to three daily reports assessing public sentiment towards it, including the supporting raw data. The same EMB contracts an advertising placement agency, which uses data to target and evaluate the reach of adverts on various channels;
- Several EMBs contract foreign companies to assure data storage and security, although data on these procurements is not always accessible;
- There has been a challenge of access to voting data by election observers and the judiciary, especially related to election petitions. For example, in the case of such a petition in one country the EMB cited geographical expanse and differences in time zones as impediments to sharing the data for the petition's process;
- One interviewee reported a case of data centres relating to elections being physically vandalised;
- Foreign contractors handling election management data do not operate in the jurisdiction of the countries where they are contracted, which could be problematic in terms of accountability as well as support services;
- EMBs lack their own data analytics capabilities. They also lack the technical expertise for public engagement on data governance in relation to elections. In South Africa, the Electoral Commission (IEC) partners with the Council for Scientific and Industrial Research (CSIR) to address data science capacity. In Kenya, gaps have often weakened responses to disinformation targeted at its EMB, and have also resulted in difficulties in technical aspect of elections management;
- There are data silos, with little cross-flow among different parts of an EMB. For example, one section may specialise in operations data such as registration of voters and parties and voting stations; another will monitor party funding. A third section will be dedicated to results and voter turnout data, including demographic information about voters in particular districts. Staff performing voter education may draw on public opinion survey data. The media liaison department may get social media data from a contracted monitoring service, and direct data from social media platform analytics (e.g. pages and posts). These cases were cited by interviewees;
- The value of different types of data varies from one stakeholder to another. Sometimes government institutions and EMBs discard certain data that may not be useful to them but still useful to other stakeholders. For example, interviewees in Kenya and Ghana were keen to get previous elections' data, and noted cases of journalists who sought election data of previous candidates, but neither could find the data that they were looking for; and
- There is some recognition of the need to change the levels of data literacy of EMB staff and to align skill sets.

EMB partnerships relevant to data (among other issues) include partnerships with media, universities, international organisations, government departments, civil society monitoring and fact-checking groups, and platform companies. This varies from one country to another. South Africa has established more partnerships with local stakeholders compared to Kenya and Ghana.

The Open Government Partnership (OGP) has been instrumental in encouraging African governments, including Kenya and Ghana, to launch open data initiatives. In Ghana, the government's open data protocol enabled various stakeholders, such as students, to access data for research purposes. From 2012 to 2016, election data was accessible through this platform, but a shift to a new system led to the discontinuation and loss of previous election data, which is now unavailable. Sustainability of data access in these cases is an issue.

Broadly, interviewees signalled a gap in EMBs' data policies, strategies, and management. There is also an absence of a holistic perspective and foresight into "what comes next" about current data acquisition and use.

Transparency is an issue with EMB data:

- The absence of data and disclosure about polling station challenges in the case of one EMB with poor communications created the space for speculation to flourish; and
- In one country, there was no data about the composition of the EMB's media monitoring committee.

Regulators and accessing data for monitoring

In parallel with EMBs, media regulators tend to have monitoring initiatives during elections that focus on traditional mainstream media (and much less so on social media monitoring). Conducting their own research directly is usually not the case, and there is a lack of capacity to interpret figures when these are supplied by contracted outsiders. The insight from one media regulator interviewed is that people are very candid on social media, making this a valuable source of raw data that should be considered.

In South Africa, the Independent Communications Authority of South Africa (ICASA) monitors mainstream broadcast media for compliance with electoral regulations. The data in this exercise arises from a team of 32 manual monitors as well as from complaints received from the public and political actors. ICASA specifically checks for compliance with regulations on allocations of airtime for political parties, and whether the content aligns with the broadcasting code of conduct. A service provider is contracted by ICASA to aggregate the data and provide visualisations of findings. Items considered to be potentially violative are assessed in more depth and referred to the regulator's Complaints and Compliance Committee. ICASA liaises with the IEC, which has a parallel process through which the authors of problematic broadcast content can ultimately be referred to the Electoral Court. Since the Advertising Research Bureau

self-regulates commercial adverts, and ICASA focuses on political adverts, the election period does not involve liaison between them. Advertising self-regulatory authorities in study countries are not involved in monitoring or hearing complaints about electoral advertising.

Privacy regulators are interested in how voters' personal information is handled by EMBs, political parties and other role-players during the electoral process. There is growing international concern about the role of data used for micro-targeting voters in campaigns. However, this research did not uncover many incidences of such phenomena at this historical point in the four study countries. On the contrary, one political party said it had previously done this, but had ceased the practice under new privacy regulations.

These data challenges faced by EMBs and other regulators in a number of countries go hand-in-hand with the wider phenomenon of low levels of public trust in statutory institutions (and also in political parties) in countries.

2.2 Observers and election support agencies

Difficulties signalled by interviewees in this category of stakeholders include the following matters:

- getting data and information from EMBs about their operations;
- EMB data issues cascading into negative impact on the databases operated by the support agencies themselves (e.g. inhibiting the potential to compare EMBs in a comprehensive way);
- limited disclosure by some EMBs beyond final results of voting;
- lack of awareness and implementation by EMBs of The Principles and Guidelines for the Use of Digital and Social Media in Elections in Africa by the Association of African Electoral Authorities ; and
- the need for support agencies to help EMBs clean and update voter registers.

Case study: Regulating data for democracy in South Africa

The South African Information Regulator combines and balances two functions: governing both access to information and data, and protecting people's personal data. This is relevant in elections in several ways:

- The Promotion of Access to Information Act (PAIA) states that public bodies (which applies to the IEC) *must* submit an annual report to the Information Regulator regarding access to information requests received and processed. In contrast, private bodies (which include political parties) are only *requested* to do so. The IEC's performance under PAIA could in future be assessed within the list of bodies which the Information Regulator identifies for detailed assessment of compliance. The Information Regulator has already surveyed political parties to establish their compliance with publishing a list of their

information holdings, although there are no penalties for non-compliant entities considered as private.

- Under the Protection of Personal Information Act (POPIA), a responsible party may not process personal information concerning the religious or philosophical beliefs, race or ethnic origin, trade union membership, political persuasion, health or sex life or biometric information of a data subject. Political parties must comply with this, although they do not have to report to the Information Regulator, nor do they do proactive disclosure of breaches of the Act. It is for the regulator to proactively decide whether to investigate, as happened in 2024 with leaked information from the IEC. In that case, the regulator has the power to issue an enforcement notice requiring the IEC to take certain steps, failing which there could be a fine.

- The Information Regulator has issued a guidance note titled "Processing of personal information of voters, and the countering of misinformation and disinformation during elections". This affirms that data from a voter cannot be supplied to a third party without the consent of the data subject.

- The Political Party Funding Act (2018), with amendments in 2024, provided for mandatory disclosure by donors and by the political recipient to the IEC of any donation above the prescribed threshold (of ZAR100 000) that has been made in a given financial year, as well as the identity of the persons or entities who made such donations. Under PAIA citizens have the right of access to public-interest records of political parties or independent candidates, including records of all donations.

2.3 Data issues and the media

The media plays a crucial role in disseminating election data as they provide timely updates on election results and voter registration, and offer an analysis of the data to provide insights into elections management. In this role, they enhance public participation in the electoral processes, and facilitate public engagement with the process beyond voting. Where multiple information and perspectives overburden the public, the media also serve to collect and assess information, sometimes in the form of data, and highlight relevant areas for focus. The media can also demystify what is contentious in the public sphere, including alerting the public to disinformation. Software expertise is needed by journalists to process voting data and link it to demographics such as age, sex and ethnicity and social science skills are needed to avoid generalising inappropriately. Current levels of skill are not conducive to many journalists seeing story possibilities in processing available raw data. Instead, reporters tend to rather use findings as presented by other institutions collecting and analysing the data. Limited capacity for visualisation of data is also a feature.

In terms of electoral coverage, interviewees and the monitoring of online webinars yielded insights into how data is important for:

- journalists attempting to verify whether certain online content is fake or not;
- journalists seeking to appraise online trolling which contests various data about the poll (although platform opacity constrains finding and assessing this data); and
- journalists assessing the impact of their outputs.

Examples where data is of value in story-discovery and story-telling:

- Journalists assess differences in data on voter registration between rural and urban areas and use this to predict electoral outcomes based on knowledge about relative party popularity in these geographies;
- Some journalists make graphics out of data to show insights such as cities where different political parties have grown vote share over previous elections, voting patterns (e.g. from Afrobarometer data), and changes in voter turnout over time. Correlating political assassinations with election years is another example of data-informed journalism;
- Journalists in one case built an elections map with profiles of candidates, as well as updates for projected winners. In several instances, interactive maps with drill-down possibilities drawing on official voting data have been created;
- The data that journalists use in election content extends beyond reporting on election data to include data informing political policy and performance issues, both past and future;

Cases where media source and use data:

- One media outlet conducted an online survey of voters' views on top electoral issues, and manually assessed the results for stories;
- Another media outlet did an online survey offering users an interactive and engaging experience, and unexpectedly found the responses to offer valuable data about their audience. The same initiative generated data showing a measurable increase in traffic to the website. Webinars by one media outlet also fed into data analytics that informed editorial decisions;
- Another media outlet partnered with a survey firm that conducted face-to-face interviews that produced data enabling the prediction of electoral outcomes; and
- In Kenya, journalists are given access to the election results Application Programme Interface (API) to enable them to report independently on the outcomes.

2.4 Fact-checkers and data

- The specific use of data-analysis techniques in electoral fact-checking is not an extensive and mainstreamed component in this realm of work;
- To identify content for fact-checking, fact-checkers receive commissions from social media platforms, and there is also some manual identification of potentially problematic content by front-end trawling of targeted accounts or hashtags. This helps to identify data sets of priority content for consideration;
- In one case, selecting what content merits attention for fact-checking is surfaced via a WhatsApp channel with several thousand subscribers. In another case, fact-checkers

themselves have joined hundreds of WhatsApp political groups, spotting falsehoods and then contributing corrections (sometimes at the expense of being expelled from the group);
- Fact-checking is constrained by the availability of reliable data to properly investigate claims such as electoral rigging, and there are fears that investigating such may lead to accusations about prejudging findings of a court. Where the EMB lacks credibility, the checkers have to fall back on data from international observers and electoral watchdogs;
- In some cases, the practical possibilities are to pre-emptively produce fact sheets about incontestable information or simply to signal when a particular claim is disputable even if its falsity is not (yet) proven;
- In one instance, there was some limited data access available to fact-checkers on the backend of the commissioning platform. This pertains to the specific content item that the company flagged for fact-checking. The data would show what the virality is, and what the view totals are;
- There is generally no comprehensive data available to fact-checkers that gives them insight into how their work resonates on the platform where dis- and misinformation is published, whether the affected content is downgraded, and what levels of "recidivism" there are for false content;
- Meta's Content Library and APIenabling direct data access is not as useful as the company's CrowdTangle (which has closed down). Another service used by some fact-checkers, NewsWhip, is said to be not as good as CrowdTangle was;
- Data from the EMB, where credible, and from the national statistics commission helps fact-checkers in the verification processes;
- Because of the cross-pollination of content between platforms, it is difficult to get data that shows the scale of virality of political dis- and misinformation; and
- While fact-checkers can reverse-image search to establish whether a media item is being used out of context with misleading significance, this is not possible with new synthetic media products such as those produced by Generative Artificial Intelligence (GenAI). In the absence of labels (whether via metadata cryptography or by visible watermarking), fact-checkers do not readily have data to assess the veracity of many of these kinds of manipulations that appear convincing and true, but which may be faked.

2.5 Civil society and data

According to interviewees:
- Civil society groups need data to advocate for voter registration, produce counter-content to dis- and misinformation, and gather evidence of risks to election integrity;
- One effort to access Meta's research interface was turned down without the reasons given;
- Some social media monitoring has to resort to OSINT (Open Source Intelligence) methods or manual scraping;
- Those NGOs with grants from donors that enable them to pay for access to social media data via brokers say that there are limitations to what is available. The data sets are only about public pages, not closed groups, and the assessment of metadata in these cases – as

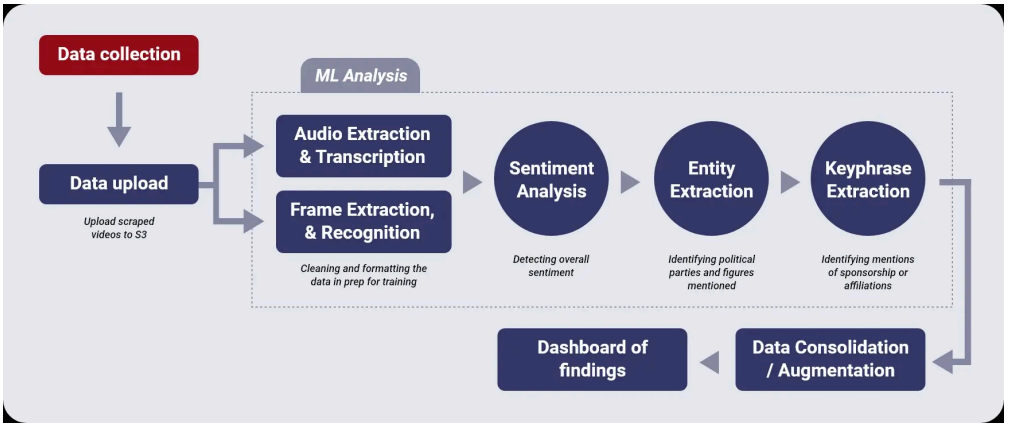
with WhatsApp – is not part of the commercial packages on offer. Labels are not part of data sets in Meta's offerings;
- Some data broker companies selling access have interfaces that enable NGOs to more easily track video content, which is complex to monitor in comparison to text;
- Donor dependence raises questions of sustainable payments for accessing data;
- Civil society groups can sometimes get insight into which political actors monitor trends because these entities then re-present these trends as evidence of (alleged) popularity or widespread sentiments;
- One NGO uses data to monitor the success of its interventions such as for assessing the reach and impact of dialogue facilitation;
- Software is used for data visualisation;
- Data analysis capacity can be a constraint, but creating partnerships can help;
- Frustration was expressed that NGO reports of abuse to platforms do not yield consistent feedback; and
- NGOs would like data to be available about official follow-up by police when reports are made to law enforcement.

Case study: Media Monitoring Africa (MMA)

MMA is a South African-based NGO that is very active in assembling and using data around electoral integrity issues. Its initiatives include:

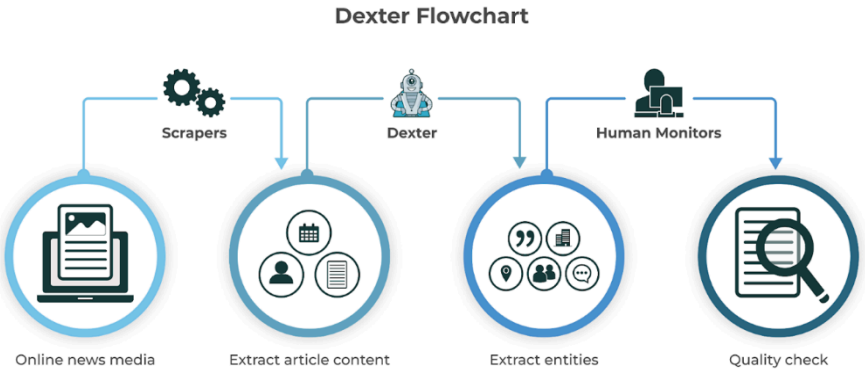
- REAL411. This is a mechanism, operated in partnership with the IEC and three online platforms. It has been used in three elections, by crowdsourcing complaints (including from a voluntary "spotters network") about online election-related content. The focus is on concerns of online dis- or misinformation, hate speech, incitement, and harassment of journalists. Nearly 290 reports were received and evaluated in 2024, covering mainly content on X, but also on TikTok and WhatsApp. Just over half were upheld by REAL411's volunteer assessment teams, and then channelled to the platforms (Meta, TikTok and YouTube, but not X) as per a memorandum of understanding with them. In 2024, MMA was a recognised flagger with Meta. A unique value of REAL411 is that it provides the public with a transparent aggregation of content problems across various platforms, thereby enabling insight into shared narratives and information operations beyond the online services separately considered. This complements the fragmented individualised reporting where users raise issues directly with platforms. However, without access to the API, the MMA could not do deep identification of disinformation networks, nor follow any further content moderation trail beyond the immediate case. According to MMA, REAL411 could be "significantly strengthened through access to information collected by the various social media platforms regarding the underlying usage data". Such access could help to identify the types of content, instigators, networks and drivers of harmful or misleading falsehoods, and to create early warning systems for future elections.

- A new MMA tool introduced in 2024 was Witi, which looks at online influencers and their stance on political parties. It is described as an experiment, using Machine Learning to produce sentiment data out of content that is manually selected from TikTok. The flowchart below shows how data is processed through Witi.



- PADRE is a one-stop political advertising library under IEC auspices, supported by MMA. The content is derived from Meta's and Google's own political ad repositories (X does not offer this facility in South Africa, while TikTok does not take political ads). The service enables voters to search for individual parties, and to see the totality of party online ads and hashtags (as distinct from what users may individually be receiving algorithmically when accessing social media). The total online ad spend on Meta and Google by political parties is also shown, as well as geotargeting breakdowns. PADRE also enables the public to match fake adverts against genuine ones via a WhatsApp interface. Parties can upload their ads but since they do not do so, MMA uses copies of ads on the Meta and Google repositories for the service.

- DEXTER is MMA's Natural Language Processing tool for a huge database of online news media articles (excluding content behind paywalls). There is a dashboard interface for checking individual titles. This was used for monitoring the 2024 South African elections coverage by the country's news media and assessing its quality, fairness and inclusiveness.



- MARS is an acronym for "Media Attack Reporting System" which collected data on 1 025 online attacks on journalists during South Africa's 2024 poll, mainly on the platform X.

2.6 Political parties and data

Interviewees in this stakeholder group said that:

- Parties are limited in their efforts to source and use data due to organisational and resource constraints. However, they noted the importance of having data sets, especially those relevant to historical voting trends and to their contemporary strategies and campaigns. A number also commission focus groups as well as larger surveys of voter attitudes and concerns, and about how their party's "brand" is viewed. Several also track social media to see what the analytics tell them, including about their rivals. Membership data is very important for planning and research. They resort to sourcing and using data from statistics agencies and the national EMB, enabling assessments of voting patterns down to ward level where possible. There are aspirations to build internal capacity to handle data, and there is use of volunteer data scientists in some cases;
- Opposition parties have concerns about how data is used for delimitations, and there are complaints about delays in releasing voter registration data which inhibits the planning of campaign trails. Where voter rolls and addresses are made available by the EMB concerned, parties use this for door-to-door canvassing (sometimes with cross-reference to data on support from earlier elections);
- Parties would like transparent reliable data on campaign spending, but there are concerns that official data on campaign spending under-reflects the realities;
- One party calibrates its spending by assessing estimated data on what its rivals spend on posters and mainstream media spots, and by examining the Meta and Google political ad repositories. There is however a lack of detailed access to data on political ad spend, with the Google and Meta repositories providing limited data sets on the targeting and placement of advertisements;
-
- Generally, the low extent of data use corresponds with parties not hiring commercial political consultancies specialising in this realm. One party says it conducts its own polling and its data team scrapes social media to pick up trends. It then uses external opinion poll data as a point of reference for its own opinion polls, with the overall reservation that sentiment analysis does not equate to actual voter turnout. The party applies the same precaution to assessing content on X, which it sees as being about echo chambers that do not reflect on-the-ground mass opinion. One party reported approaching outside bodies to help with data mining (especially about the virality of their own posts on social media), but it found these entities were lacking in software and capacity;
- While parties have social media strategies, only some of these are data-driven. Privacy laws prevent them from using data for micro-targeting. Parties get some data directly from social media companies about the performance of their own ads and their political pages; and

- Political advertising does not especially appear to be data-driven, especially when done directly rather than through ad agencies.

2.7 Researchers

Interviews with political scientists elicited these insights:

- Generally, academics say they rely more on secondary data sources, especially opinion poll survey data which they use for predicting electoral outcomes. Surveys are very expensive to conduct, and very few of the academics studying elections collect first-hand data. Some have partnerships with big political parties which can give privileged access to deep data on voter attitudes from research that is commissioned by these parties;
- There is caution that opinion polls as data sources need a lot of evaluation in terms of the methodologies used, including whether they were conducted by phone or face-to-face or another method used the survey structure, wording and sequence of questions; the margins-of-error and sampling representativity employed; considerations in terms of language the aggregation software used; and how the findings were extrapolated to the wider population universe. This is necessary when comparing different findings between polls. Political scientists also highlight the need to look at changes and trends over time, as well as other information cues, when interpreting the data. There is limited cross-country comparison;
- Another data source for election analysts is the EMB's data, especially on voter registration and turnout. This is combined with official population statistics data (which EMBs use as a basis to calculate their figures);
- Work with statisticians is rare, although some academics do tabulations and use spreadsheets or SPSS packages. Visualisation of findings is limited;
- The interviewees say they tend to interpret findings qualitatively rather than do digital analytics on the data. There are relatively few African political analysts trained in quantitative research methods, such as regression analysis and pattern detection, although there is an acknowledgement that more data-oriented capacities need to be built. The use of AI in data analysis to date has not proved very useful; and
- Overall academic researchers state they have very little support capacity in conducting electoral research.

Among political scientists, there is mixed interest in social media as a data source. One interviewee warns that social media is not representative of wider trends. Another argues that social media does not play an independent role in African countries and does not shape voting patterns in the way it might elsewhere. Mainstream media are regarded as more influential, although there is little in the way of getting data on this realm to assess this as an electoral factor. The argument is made that any social media findings must be rebalanced since the people and the behaviours online are not representative of the actual voting populations. With such caveats, it is nevertheless recognised that the real-time possibilities from live data analysis on social media could give some insights. The

point is made by interviewees that it is important to triangulate social media data and findings with survey data if there is to be an assessment of online attitudinal trends about whether to vote and/or whom to vote for.

Further points emerged from a meeting of ten research groups (some university-based, others in NGOs) in regard specifically to data access and analysis of online content about the 2024 South African election:

- They issued a public statement urging platforms to provide data access, pointing out that: "[R]esearch sampling of a selection of content when conducted from the outside is limited in terms of identifying co-ordinated flows of potentially harmful content. The alternative of accessing platforms' data sets through commercial providers is very costly, and also offers a rather limited range of data". The groups with grant funding largely resorted to expensive data brokers, reinforcing a disproportionate focus on X as a platform. Hashtags were used to identify narratives, and content was tracked in several languages;
- There was limited checking for cross-platform coordination of potentially harmful content. Few of the groups use alternative research methods to triangulate their findings, although one did conduct online surveys to give insight into perceptions about the credibility of disinformation narratives, while another examined mainstream media narratives which enabled comparison with those on social media. Online impact was assessed by looking at impressions, follower numbers, and engagements with content, though offline research was not done to compare with real-world impact;
- Doxing, political ads and (paid) influencers were not widely monitored. Little attention is given to images and audio, and also rare was assessment of synthetic content and use of hyperlinks. Analysis of followers and engagements yielded some insight into networks;
- Through using purchased data opportunities and deploying data science capacity, researchers were able to identify "influence for hire" on X, with key accounts (often pseudonymous) of "mega influencers" being used to "hyper-amplify" messages through smaller accounts of paid "nano-influencers". This found that mutual "follow trains" were being fostered, while bots were used to artificially inflate follower numbers, with WhatsApp groups being created to further coordinate messaging. The Institute of Security Studies' report on the 2022 Kenyan elections found through online data analysis and offline interviews that candidates' election discussions were significantly disseminated through paid influencers;
- An important gap identified was that platform policies were not referenced as benchmarks, and platform performance in dealing with potentially harmful content was not assessed; and
- Relations with the media to publicise the research results were patchy, with just some of the research groups having concrete plans to disseminate their research findings to the broader public.

Case study: Council of Scientific and Industrial Research (CSIR)

The CSIR in South Africa uses live data from voting returns in clusters of voting stations, to be able to predict real-time trends and eventual electoral outcomes. Their model of statistical clustering and mathematical algorithms achieves good predictions from a small sample of results. The CSIR has a partnership with the public broadcaster, the SABC, where the process is part of the live coverage on election day. The Council also monitored almost three million posts on X during the 2024 elections in South Africa, showing that 42% were negative in sentiment and only 14% positive. Whereas in 2019 the CSIR could automate its data collection on X, the closure by the platform of its free API access meant that a more tedious method was needed in 2024. According to a presentation from CSIR in 2024: "Researchers need assistance from government organisations in expediting data access from social media platforms."

2.8 Platforms and data

Attempts to contact Google for this report did not yield results. From information and links provided by TikTok and Meta (for Facebook and Instagram) it is evident that data is relevant to their election-relevant operations, from content moderation to the screening and placing of advertising. However, data sharing as an issue seems absent from platforms' engagements with fact-checkers, the media and EMBs.

To the extent that Meta's AI identifies content that violates terms of service, it is trained on structured data samples and treats new posts as data for processing by automated content moderation algorithms. Processing involves predicting whether a particular item is potentially violative and whether it is part of viral disinformation. When so identified, a set of enforcement technologies then determines whether to take an action, "such as deleting, demoting or sending the content to a human review team". Decisions by human reviewers (both full-time employees and outsourced sub-contractees) are then said to interpret the flagged content in context, and this adjudication is fed back into further ongoing training of the technology. In 2023, Meta stated that it had 40 000 people worldwide working on "safety and security for global elections", in a reference that may have included commissioned fact-checking services. The company says it removes content that interferes with voting or calls for electoral violence, thus reducing distribution of "false news" about elections. According to Meta, security teams investigate and take down coordinated networks of inauthentic accounts engaged in influence operations, with some experiences between 2017 and 2021 as regards the four African countries in this research available on Meta's website, although granular data is not supplied.

Meta also says it provides details on registration through a Voting Information Centre and through "in-app" notifications. Examples are only given for the United States. Regarding advertising, Meta states that political advertisers have to be verified, and that there is transparency on owners of political pages and groups. It flags its archive of political and issue ads showing who is targeted and who is paying.

According to Meta, it set up a South Africa-specific Elections Operations Center to identify potential threats and put mitigations in place in real-time. It also had a fact-checking network in five languages. The company said that "[w]hen content is debunked by these fact-checkers, we attach warning labels to the content and reduce its distribution". The company also said that fact-checkers (Africa Check and AFP) had access to its Meta Content Library. However, data sets are not made available that could help assess the company's performance in acting on electoral threats and in attaching warning labels. This data is also not in the research fields on the Meta Content Library. The company further announced that it has set up an "Election Center" for the South African poll, with advice to election contenders on verifying and securing accounts and reaching voters (e.g. via the WhatsApp Business App, which includes advertising performance for purchasers of this service).

For its part, TikTok also announced an in-app election centre, which carried information from the IEC, and which the company said would be accessible through search or clicking on labels on election-related content. The company further noted a partnership with the South African Human Rights Commission and pledged to "label election-related content, unverified election claims, and state-affiliated media accounts". There is no detailed indication of how such steps were to be implemented, monitored and reported upon.

Google, Meta, and TikTok (but not X) had a Framework of Cooperation signed with the IEC and MMA through which the companies committed to addressing reports of electoral disinformation (see REAL411 write-up inMMA case study above).

Experiences by civil society suggest potential shortfalls in the platforms' performance in adhering to their public announcements. Three invitations by editors to the platforms to dialogue on risks and mitigations about online threats to the South African elections were generally ignored. Google did release a range of globally generic information, as did Meta and TikTok. These were often without meaningful data as to what provisions would be applied in South Africa (e.g. specific technical and human capacities in local languages for electoral content moderation), how threat detection would be done, and how expressions about advancing authoritative information sources and voter education would be promoted in practice. At the time of writing (March 2025), a data and knowledge gap remains about if and how the digital platforms assessed their performance in regard to election integrity.

Two South African NGOs then made formal requests under the country's PAIA for records showing the platforms' electoral plans. This was met by responses that these entities did not fall under South African jurisdiction. However, the Information Regulator rejected this in August 2024 (in regard to Meta) and initiated discussions with the platforms concerned, who in 2025 maintained their argument on jurisdiction. The matter continued in March 2025.

Independent assessments, largely using social media data bought via commercial brokers, showed a range of content that can be seen as clearly breaching the platforms' guidelines and enforcement efforts. This included allegations of the election results being fraudulent ("the big lie"), as well as polarising racial and populist content, and content that sought to stoke ethnic tensions. Russian

influence operations on X, the work of paid influencers, and photos and videos presented out of context were detected, despite the platforms' commitments. AI deepfakes were not extensive, and were partly addressed by platforms.

On digital advertising, the Legal Resources Centre and Global Witness submitted ten (dummy) hate adverts targeting women journalists, in four languages, in the lead up to the South African election, finding that such paid-for content would generally be accepted by Facebook, TikTok, YouTube and X.

There is however not evidence of detailed data being collected and analysed on the content of published adverts.

Where big data analysis was referenced, this showed that notwithstanding platform policies, there is an unabated epidemic of online violence against women journalists that was of high relevance to information integrity and free expression during the South African election. Without more comprehensive access to data, it is not possible to know how this phenomenon relates to platforms' election threat analyses and mitigation measures.

3. Conclusions

What this report shows is that particular attention is merited for the specifics of data regarding the key moment in democratic life which is elections. At the same time, as one interviewee pointed out, if voters are to make informed choices, they need meaningful data about political parties' policies and track records. This wider perspective points to the way that the issues raised in this study could be helpfully considered alongside the role of data in society generally. The governance and use of data and access to it during elections is distinctive, but it is also part of the bigger policy and regulatory picture about data in society at large.

Challenges surfaced by this research include the following:

1. There is uneven awareness of the range of roles of data in elections, although there is also some recognition of the importance of coming to grips with this;
2. There is some controversy around the protection of personal information and accuracy issues with respect to elections data. Especially in contexts where this controversy exists, data is far from being treated as a neutral resource, but instead is harnessed for biased political agendas;
3. Structural, technical and cost hurdles inhibit stakeholders from accessing pertinent public and private data;
4. The opacity of platforms underlines the significance of increasing calls to open up access to granular and comprehensive data for the purpose of monitoring African elections;
5. Limited capacities affect stakeholders' abilities to assess and process data; and
6. Uneven access to data and capacities for the use of data weaken efforts to identify and counter threats to information integrity and to assess what the platforms are doing about this.

Recommendations

1. General:
    a. Roundtables of the relevant stakeholders for African elections could overcome low awareness and fragmentation amongst them, and highlight the significance of data issues and access to data for these key moments in democracy;
    b. To monitor changes over time, ongoing research is needed to assess the datafication of African elections;
    c. To address gaps revealed in this report, strategies are needed by EMBs and data/information regulators to ensure there is adequate governance of data availability and privacy in the interests of election integrity; and
    d. Wider national policies and African standards should be considered when developing national approaches,thereby helping to ensure optimum approaches by stakeholders to the use of data in elections.
2. Supply-side:
    a. EMBs need to protect personal data but also to allow granular access to their own data;
    b. Comprehensive data on campaign spending by political parties is needed; and
    c. Platforms need to provide comprehensive data access to researchers, fact-checkers and journalists for monitoring electoral online content and behaviours, as well as for assessing the performance of platforms and how this impacts election integrity.
3. Demand-side:
    a. Efforts are needed to increase data awareness and skills in most stakeholder groups; and
    b. Partnerships and cooperation should be enhanced.
4. Governance issues:
    a. The AU's instruments on elections, data and AI merit revisiting to assess fitness for purpose and to advance implementation at country levels. There is an opportunity to do so in the ACHPR mandate of the Special Rapporteur to investigate data issues;
    b. States and EMBs can do well to revisit the adequacy of national policies, regulations and strategies that encompass data and elections;
    c. As has been proposed elsewhere: "Regulators are encouraged to actively pursue cross-regulatory cooperation across electoral, human rights, privacy, and other regulatory spaces as these expectations extend to registered political parties, campaign organisations, commercial data brokers, analytics firms, advertisers, and social media platforms";[5]
    d. States should consider convening multistakeholder human rights impact assessment exercises ahead of elections. These should consider potential data benefits and harms, and related risk and mitigation opportunities, and also

[5] Philippe Dufresne, Privacy Commissioner of Canada and Chair of the Data Protection and Other Rights and Freedoms Working Group of the Global Privacy Assembly, and Ana Brian Nougrères, United Nations Special Rapporteur on the Right to Privacy, https://inforegulator.org.za/wp-content/uploads/2020/07/Statement-on-privacy-and-democratic-rights.pdf

commission transparent evaluations of data-related issues after a polling periods; and

e. Data relevant to the creation and circulation of GenAI artefacts depends on the cooperation of users, political contenders, AI companies, and platform companies. Norms and rules in this area, pertaining to electoral content, are important for EMBs in particular to develop.

Appendix A: Background and framework of this report

Since the Cambridge Analytica scandal, there has been increased global awareness about the role of data in elections. In the Global North, there is a growing amount of literature on the issue, as well as extensive news reporting on what are being referred to as “data-driven” elections.

Within this context, the datafication of elections presents both unprecedented opportunities and complex challenges for democracies in Africa. While data-driven technologies have the potential to enhance transparency, civic engagement and electoral integrity, they also raise concerns regarding dis- and misinformation, voter manipulation, and infringements on human rights. These challenges have significant implications.

At the heart of tackling this solution is the question of access to and use of data. This is relevant to EMBs, which depend on data inputs for their work, and which also output data to the general public. Political parties are also major stakeholders regarding a range of historical and live data that is used in their political campaigns and in their subsequent roles in government and as the opposition. A lack of access to data also limits the ability of media and researchers to effectively deliver on their roles in monitoring voting elections, combatting dis- and misinformation, and fostering participation and public trust in elections. Access to disaggregated data on voter turnout could enable civil society groups to identify patterns of voter suppression or apathy and advocate for more inclusive electoral processes. Similarly, access to campaign finance data can help journalists track the influence of money in politics and hold candidates accountable for their campaign promises. While many African countries have demonstrated a commitment to access to information through legislation, the concept of data as a public good and as part of the fundamental right to access information, particularly within the electoral process, remains underdeveloped.

The challenge of data access extends beyond national borders. The dominance of international tech companies in shaping online political discourse, and their role in holding this data in their own silos necessitates a framework that addresses the transboundary flow of data as well as public interest access to private data holdings. The European Union's Digital Services Act, which mandates data sharing by social media companies for research purposes, highlights the growing global recognition of access to data held by tech platforms as a public good. Platform companies already share their data through formal arrangements between them and researchers in the Global North, while similar transparency initiatives between the companies and researchers in Africa do not exist (beyond in-principle access akin to other countries to some of Meta’s data). This raises concerns about unequal policy implementation and potential data colonialism. Without mechanisms for accessing data held by these companies, African stakeholders will be left ill-equipped to address the growing threats of datafication and, at the same time, miss out on its opportunities.

Attempts to frame the issues in Africa can draw upon a number of instruments, albeit they do not delve deeply into data issues. These include:

- The 2017 Guidelines on Access to Information and Elections in Africa;

- The 2019 study on Proactive Disclosure of Information and Elections in South Africa by the Centre for Human Rights, University of Pretoria;
- The 2023 Africa Union Data Policy Framework;
- The 2024 Principles And Guidelines For The Use Of Digital And Social Media In Elections In Africa; and
- The 2024 Resolution on promoting and harnessing data access as a tool for advancing human rights and sustainable development in the digital age(ACHPR/res.620 (lxxxi))

This research builds on work by RIA that gauges the appetite of African researchers for data held by both the public sector and tech intermediaries, and the responses to this. Beginning in November 2023, and in partnership with International Media Support, RIA brought together actors interested in the nexus of technology, elections and media. A public statement from this engagement involving 50 stakeholders from 11 African countries called for platforms to provide Africans access to data on a par with other regions. This in order to monitor and formulate joint mitigation responses to disinformation and hate speech given elections in nearly 20 countries in Africa over the course of 2024. One follow-up meeting, focused on the South African elections and convened by RIA and Data Science For Social Impact (DSFSI) at the University of Pretoria, made a similar call. This pointed out that without real-time collection of data at the source, external researchers are hamstrung in monitoring election threats online.

This study complements work by the African Alliance for Access to Data, of which RIA is a founder member, to develop a tool for EMBs, political parties and civil society, in order to get to grips with data issues. It also synergises with the Association of African Electoral Authorities guidelines mentioned above for social media in elections that include reference to access to data.

This research is also part of a broader IDRC project which revolves around advancing democratic, inclusive, and equitable governance in Africa through the use of data and digital technologies. The project aims to strengthen the implementation of the AUDPF by focusing on public digital infrastructures, data categorisation for cross-border data flows and national governance, and institutional strengthening.

The outputs of this research are intended to contribute to standards on data in this area, as well as access to quality data which is relevant more broadly, such as for building African data commons and advancing open access on the continent. The findings will also be useful for capacity-building efforts around this focus area.

Preliminary conceptual framework

Local interests/local data holders	**Foreign interests and data holders**
Election bodies and observers: • data that can help inform voter education; • data relevant to countering voter suppression;	Social media, search companies and GenAI companies:

<table>
<tr><td colspan="2">data relevant to threats and mitigations about electoral rights-related content;
voter registration data;
Vote counting data; and
advertising spending data.</td><td colspan="2" rowspan="4">data on information exchange, including political advertising, influencers and groups;
hate speech and dis- and misinformation networks;
data on political engagement;
data on political preferences; and
data on political violence against women candidates and journalists.</td></tr>
<tr><td colspan="2">Political parties:
data about voter sentiment trends;
data about party online performance in terms of impressions and reach; and
data about advertising spending.</td></tr>
<tr><td colspan="2">Telecommunications companies providing internet access services and others in the tech stack (e.g. AI, device makers, mobile operating systems) and infrastructure for access to data and information:
data about traffic and formats; and
data security.</td></tr>
<tr><td colspan="2">Civil society:
data about online threats to electoral integrity; and
data about platform mitigations.</td></tr>
<tr><td colspan="4">↑↑↑
Supply-side</td></tr>
<tr><td colspan="4">Data categories</td></tr>
<tr><td>Public data</td><td>Private data to be shared only with verified stakeholders in the ecosystem</td><td>Private data to be shared and agreed to be shared by specific stakeholders</td><td>Private data only shared within the institutions</td></tr>
<tr><td colspan="4">Demand-side
↓↓↓</td></tr>
<tr><td>Costs</td><td>Access infrastructure</td><td>Capacity of use</td><td>Minimisation of harms</td></tr>
<tr><td colspan="4">Opportunities for creating shared data ecosystem</td></tr>
</table>

Appendix B: Methodology, interviewees and sources consulted

Interview questions:

The structure informing the interviews was to investigate perceptions of data and elections on both the "demand" and the "supply" side. The topics covered included:

- Awareness of data/information around elections and usefulness of this data/information to the mission of the given stakeholder group;
- Surfacing knowledge and learning from experiences (existing partnerships, good practices);
- Access challenges;
- Capacity challenges;
- The stakeholders' own data (where applicable), their willingness to share this data, and their preferred governance structure for the sharing of data; and
- Whether they used visualisation to communicate the value of assessed data.

Question bank used for interviews:

1. Your awareness of who the primary users of data during elections are (e.g. parties and PR campaign agencies, platforms, EMBs, monitoring groups including fact-checkers, media and journalists, pollsters, academics, state surveillance agencies, the business sector, voters);
2. Current usefulness of data to your mission: What data do you use and how do you use this data? What capacity and tools do you have for analysis and visualisation?
3. Who has data that you could use?
4. What data access challenges do you have?
5. What are your experiences of data-sharing partnerships and good practices?
6. If data were to be shared, including your own data (where applicable) and assuming a willingness to share, under what preferred governance structure would this sharing take place?

Persons consulted for this research:

The authors would like to thank the following for their generous insights, while affirming that the content in this report is their own responsibility:

- Akinduro, Olufunto. International Institute for Democracy and Electoral Assistance (International IDEA)
- Bapela, Kate. Electoral Commission (South Africa)
- Bird, William. Media Monitoring Africa
- Booysen, Susan. Mapungubwe Institute for Strategic Reflection
- Chaniwa, Farisai. Media Monitors
- Chetty, Yossabel. Centre for Analytics and Behaviour Change
- Chinaka, Chris. Zimcheck
- Clifford, Cayley. Africa Check
- Donkor, Wisdom. Open Data Hub
- Duvenage, Andre. North West University

- Gerenge, Robert. UN Development Programme
- Guta, Chengetai. Movement for Democratic Change
- Kgatse, Mamedupe. Independent Communications Authority of South Africa
- Kotze, Dirk. University of South Africa
- Mashigo, Busisiwe. Independent Communications Authority of South Africa
- Meta representative
- Modisane, Hilda. Executive Secretary, The Electoral Commissions Forum of SADC Countries
- Mokoena, Lefa. Independent Communications Authority of South Africa
- Moyo, Tabani. Media Institute of Southern Africa
- Ndebele, Zenzele.Centre for Innovation and Technology, CITE
- Otieno, Churchill. Kenya Editors' Guild
- Penplusbytes
- Phiri, Godwin. Zimbabwe Media Commission
- Pollicy
- Rapoo, Thabo. Electoral Commission (South Africa)
- Ruhanya, Pedzisai. Media Institute of Southern Africa
- Sarupen, Ashor. Democratic Alliance
- Shulz-Herzenberg, Collette. University of Stellenbosch
- Smith, Scott Peter. Mail & Guardian
- TikTok representative
- Tlakula, Pansy. Information Commission (South Africa)
- Wamugu, Eric. Code for Africa

The authors would also like to thank Malvern Mkudu, from Misa Zimbabwe, for assisting in identifying interviewees.

A consultation around these issues at a session in the African School of Internet Governance took place with African parliamentarians in November 2024. They highlighted the challenges of internet access limitations for elections, as well as violent extremism in the Sahel, and a lack of transparency to tackle fraud where states and service providers were complicit. The issue of data sovereignty and countries being able to regulate their own data was also raised.

Additional sources:

CSIR. Unlocking insights: Social media analysis during elections. Presentation at GCIS Digital Media Forum (25 July 2024)

Dufresne, P and Nougrères, AB. 2023. Joint Statement on Privacy and Democratic Rights.

Findlay, K. 2024. Influence operations in South Africa's 2024 National Elections

International News Media Association. 2024. Webinar: How South Africa's newsrooms covered historical polls in a digital-first news environment

Media Monitoring Africa. 2024. Media Performance Review. National and provincial elections 2024.

Ngamita, R. 2024. It Is becoming Impossible to do Internet Research

South Africa: Promotion of Access to Information Amendment Act, Act 31 of 2019

South Africa: Political Party Funding Act, Act 6 of 2018