

Towards an African alliance for meaningful access to Intermediaries' data holdings



Workshop17
17 Dock Road
V&A Waterfront
Cape Town, South Africa
Phone: +27 21 447 6332
www.researchictafrica.net

1. Background

Research ICT Africa (RIA) is a think-tank with a track-record in generating knowledge on digital matters. This report covers a project supported by the Action Coalition for Meaningful Transparency (ACT), of which RIA is a member. The topic is platform transparency and data access for African stakeholders. The findings arise from desk research, key informant interviews (Appendix A), as well as consultations and discussions at four events around the continent (Appendix B).

2. Introduction

Access to digital data, and the use of such data, is a significant variable for progress in Africa towards the sustainable development goals and Agenda 2063. This is evidenced by advances such as the rise of generative AI, which is trained on large data sets.¹ An illustration of the possibilities was already highlighted back in 2013 in a report [by USAID and UN Global Pulse](#) on Kenya which stated how useful it would be to stakeholders if there were greater access to social media and telecoms data.

Critical to any access to third-party data is both an adherence to the principle, but also the existence of optimum mechanisms to mediate data access. This is important to prevent untoward corporate controls on the one hand, and Cambridge-Analytica-style privacy abuses on the other. Yet the two imperatives - privacy and access objectives – are not inherently incompatible. A balance can be struck in line with the respective human rights involved in each aspect.

The access issue is not without debates about [geopolitical and neo-colonial](#) issues. At the same time, these controversies only highlight the value for Africans in the [many different sources of data](#) across specific areas of life on the continent. In this context, among the biggest holders of African-generated data at scale and over substantial time periods are electronic communications Intermediaries – i.e. social media platforms (and to a lesser extent, telecommunications (telco) and satellite companies). Comparatively, most African states are gravely lagging in both datafication and access issues.

It is the social media platforms that constitute the main focus of this report. Many of them are members of the [Global Network Initiative](#) and have made general commitments to transparency and even taken some steps towards openness in some regions. African interests in their data holdings cover - but also go beyond - what is currently availed in transparency reports or [existing access possibilities](#) as currently (and differentially) afforded by the companies (see Appendix C). Meanwhile, the predominant African reality has been one of limited access to and uptake of social media data assets. This mirrors African stakeholders' difficulties in accessing (and processing) public data in general, as well as data from other private sector actors. Advances in regard to social media platforms could stimulate progress in these other areas and aspects.

Research underpinning this document² reveals an African appetite for increased access and use of social media platform data. But it also records perspectives that there is low awareness in the wider society of the significance of data, and it shows gaps in data literacy and governance even amongst

¹ This report aligns with the perspective of the data under focus as being those “which are collected, extracted, gathered, or scraped from a web-based platform such as a website, social networking site, mobile application, or another online space” ([The state of digital media data research 2023](#)).

² Key Informant Interviews with 35 experts and data-related practitioners in diverse sectors from around sub-Saharan Africa, and views expressed by participants in four events convened in the framework of this project.

interested stakeholders. Indicative of marginalisation from a global point is a [survey of 32 public-oriented organisations](#) analysing disinformation on social media and which reflects no African groups, thus signalling the scarcity of such initiatives on the continent. Nevertheless, there is indeed interest amongst academics (particularly those advocating Open Science), as well as data journalists, NGOs with interests in disinformation and in other subjects (eg. gender equality, climate-related online content, fact checking). There is also interest amongst regulatory instances (eg. Electoral management bodies) and data-based start-ups.

Currently, access to platforms' data holdings is largely achieved by going through data brokers (which are too expensive for many African stakeholders) and/or to informal and limited scraping of public-facing data on the platforms – rather than direct access to these companies' Application Programme Interfaces (APIs). An alliance of interest groups could make advances in opening up the situation for public interest purposes. A [specific and modular focus](#), such as around threats to the integrity of elections (as a major national public interest moment in countries affected), could catalyse better data access and public interest usage.

3. The value of access to data in Africa

In the words of one key informant, “very few Africans have conceptualised data as a critical cog for development”. However, the African Union (AU) in 2022 has agreed on a [Data Policy Framework](#), with an implementation plan developed in 2023. This framework received attention at the 2023 [African School of Internet Governance](#). The AU initiative highlights the value of data in general to researchers, policy makers, NGOs, the technical community, branches of government, IT start-ups, NGOs, etc. It does not elaborate on data held by private entities, but it does underline that data can be an essential component of evidence-based development policy and practice in Africa.

Of relevance to the issue of accessing platform data, African countries have made strides towards improving broad policy ecosystems about data. By mid-2023, the Malabo convention on cybersecurity [had entered into force](#) with the required minimum of 15 states ratifying the agreement. [Researchers assess](#) that of 55 African countries, 35 have comprehensive data protection laws and three have draft laws. A total of [27 states have access-to-information laws, with 17 draft laws in existence](#) - some of which extend to private sector data holdings. The African Union Commission's [Declaration of Principles for Freedom of Expression in Africa](#) (2019) notes that “Every person has the right to access information held by public bodies and relevant private bodies expeditiously and inexpensively”, and that “Every person has the right to access information of private bodies that may assist in the exercise or protection of any right”. Civil society initiatives reflecting on access to data include the June 2023 [“Outcome Statement from the Regional Conference on Information and Communication Rights in Africa”](#).

As UN members, African countries have also signed up to the Sustainable Development Goals (SDGs), where indicator 16.10 calls for “public access to information and fundamental freedoms” and reporting thereon.

Data collected and held by public entities in Africa became more available during the Covid-19 pandemic, providing transparency to publics (and an evidence-based checking) about what underpinned official positions. The argument of one informant interviewed for this report is that the case for data needs to be shown at local level, for example with market vendors seeing the practical value of data about online orders during the Covid lockdown which had shut down physical trading venues.

The Open Science and Open Data movements have a presence in Africa, while data-journalism and NGOs monitoring online hate and disinformation are hungry for platform data access. Start-ups in the AI space have interests in platform data for training large language models for under-served African languages. Regulators, planners and officials could profitably use telco ISP data records for administration and service delivery. A handful of academics are engaged in training programmes for digital research methods, but lack live data sets in both their training and in applications of these new skills.

There is still a lot of progress to make in raising wider awareness, including with policy makers, about the value of access to private data holdings. But prospects exist for this to become an issue on the African agenda – and the obstacles (see below) can be addressed.

4. Barriers to availing and accessing platform data in Africa

The Action Coalition on Meaningful Transparency (ACT) has categorised 32 possible barriers to platforms providing greater data transparency. These cover companies' perceived or actual risks that can hinder or prevent the disclosure of information. The research for this report builds on the ACT model, but expands to consider obstacles on both the supply side (the platforms) and the demand side (current and would-be users of platform data).

Informants for this research were asked how they perceive platforms' assessment of barriers to data access. These are the issues discussed with them:

- a. Platforms' policies/ unequal policies/ lack of policies
- b. Legal (eg. over-determining data protection regimes)
- c. Technical hurdles
- d. Commercial aspects
- e. Platforms' calculus about harms and mitigations

The informants were also asked about what barriers they perceive on the demand side. These included the fields of:

- a. State of awareness amongst stakeholders
- b. Concerns about legality of informal/unauthorised access and use
- c. Challenges of skills and technology
- d. Economic barriers (costs of access and use)
- e. Mitigations/work arounds.

4.1 Obstacles on the supply side of platform data

Overarching responses from key informants are that the primary obstacle on the supply side consists of platforms' policies and practices which disempower Africans in comparison to how other regions are treated in terms of access.

One informant notes: “If we don’t ask platforms for data, we won’t receive”. In a similar spirit, another informant puts the onus on the “demand” side: “Whether big tech or elsewhere in the private sector: you come up against a barrier and you have to dialogue, build trust and persuade them of the benefits of sharing data”. However, many more informants point to the difficulty of identifying and accessing the right interlocutors at decision-making level in the companies with whom meaningful discussions can occur. This concern was also shared as a main concern among civil society participants during our consultations at the Freedom of the Internet Forum Africa 2013 (FIFA). One informant who has a framework agreement for co-operation with several platforms around election monitoring says their modus operandi is to keep data confidential.

Informants widely believe that if platforms have the will to share data, these multinational companies could easily comply with legal and regulatory environments in Africa. A case was cited of [Google and AirBNB sharing data with South African Tourism](#), without legal hurdles blocking that relationship.

Regarding the technical level of data access, respondents indicate that platforms have created API access and vetting systems for researchers elsewhere in the world, meaning therefore that there should not be major effort needed to extend these to Africans.

On commercial issues, informants note that while private brokers using platform data are successful in selling data access on the continent, platforms providing direct complementary or reduced rate access to African researchers for public interest purposes are unlikely to harm commercial markets for data access in most African countries.

Likewise, a number of informants dismiss the view that intermediaries may also be concerned with what they see as giving away a monetary asset which could advantage commercial rivals. The informants believe that such competition is very far from being a reality within the African market. Instead, they note that data partnerships that avail public and private data more widely, could end up being of most advantage to the corporates. An example is start-ups adding value not only to their own services but also to the foundation models of big data-holders within the wider ecosystem.

Further on this theme, one informant says that collaborations often see tech companies wanting “data in”, but not out. In this perspective, in regard to donations to African governments, start-ups and civil society to create and release open data, big tech stands to be the biggest beneficiary. “The amount they will gain is multiple factors more than what local people will gain. They all want you to use and add to their platforms, and they want to build their moat.” Overall, however, the concern is not per se uneven benefits from wider data sharing, but inequalities in terms of who actually contributes data and in the magnitude of differential benefits from a common pool.

It is perceived by many informants that the companies’ conceptions of potential harms of data disclosure on the continent are overblown. They argue instead that there is not sufficient moral incentive for the platforms to implement mitigations for what harms might materialise. The absence of human rights impact assessments by platforms means there is no real basis or threat analysis of increasing data availability in the context of African countries and issues.

Respondents recognise platforms' corporate concerns about possible damage to their public image, and they are aware of critical journalism or NGO reports arising from data analysis.³ They counter that there are also adverse reputational signals at stake if intermediary companies simply cover up by flatly refusing access. The informants assess that such a stance would demonstrate that platforms are set on different standards for data access in Africa as compared to other world regions.

4.2 On the demand side for platform data

Informant responses in general cover the following points.

It is almost universally noted that there is insufficient awareness of the issue in Africa, coupled with low levels of skill even amongst interested parties. Plainly put by one informant: "We need to build a culture of appreciation of data", with associated national data strategies to access and use data. The same person echoes others in observing: "We are not even using the data we already have, and we need to address challenges of capacity alongside access".

Another informant says the primary problem is not a lack of access to platform data: "It is that we are not building capacities to take advantage of a situation that is emerging, there is a huge space, but governments are not investing in this'. In his view, there is a need to build demand for data at grassroots level. One idea is for universities to be funded based on their results in supporting people through adding data to the value chain. Another informant says: "Do our local policies make sense to have access? We have to battle even at our university to promote data access. Conscientisation is needed."

In terms of legal barriers, some informants were asked about moves such as by data-protection authorities in Europe calling on companies [to prevent data-scraping](#), and X's legal actions against an NGO for allegedly contravening anti-scraping terms of service⁴, chills informal access arrangements. Most informants doing Open Source Intelligence (OSINT), interviewed for this research, did not anticipate legal challenges, but there were some who feared that they could become caught in this net. Some informants also perceive a potential risk that African researchers who use big data in ways that embarrass powerful interests, may attract retaliation. They call for protections needed such as exemptions from overly-broad laws on "cybercrime" and copyright.

Even if the hurdle of access to platform data is surmounted, there are significant challenges of still cleaning the data and of access to computational requirements (including software) to process it. In addition, there are shortages of data analytical skills. These include limited expertise in understanding digital research questions and methods, and a shortage of skills in data science, as well as challenges in access to software and processing power. The [African Open Science Platform](#) and a number of [smaller initiatives](#) are seeking to mitigate these matters, according to some informants. An informant working on data governance and digital rights says that lack of data analysis skills is a huge problem to civil society's use of data to advocate for digital rights. As they tried to find out how digital civic rights organisations were affected by Twitter's (now X)

³ Journalistic [coverage reflecting badly on Meta](#), using the company's Crowdtangle interface, [is reported as leading Nick Clegg, Meta vice president of global affairs, to complain](#) that "our own tools are helping journos to consolidate the wrong narrative."

⁴ In mid-2023, X [sued the Center for Countering Digital Hate](#) for a fall-off in advertising. It said the Center "[as a registered user of X, scraped data from X's platform in violation of the express terms of its agreement with X Corp](#)"

commercialisation of its API data, they realised that the civil society organisations in the region have not been using the Twitter API data in the first place. ‘They didn't even know what APIs are.’

There are concerns amongst some interviewees around what kinds of software, especially free and open source, can best serve African users, and about the issues of sharing data with cloud-based services. In several cases, resources have had to be found to build analytical tools customised for the users. The potential use of AI to access and process data is debated.

Access to a level of data holdings (with some Intermediaries) by universities’ entities as well as NGOs is happening via brokers like [Brandwatch](#), [Meltwater](#) and [Netbase](#).⁵ However, this exceeds the budgets of many would-be data users. It is also remarked that large sums are paid for what one informant calls “hobbled access” to commercial brokers with constrained interfaces, whereas there is interest in raw data per se, and in a much wider range of data beyond that on offer. Additionally, there are concerns about the inflexibility of paid access packages.

In terms of mitigations about lack of affordable access, some informants say they rely on informal scraping of public-facing data, or on bypassing these data holders by collecting their own data (eg. from government pdfs or from crowdsourcing public data). Given the small scale of operation and public interest ethics at work, most informants feel that Africans are not especially vulnerable to legal actions by either platforms or data protection agencies.

Informants show understanding that not everyone should have access to platform data. They advocate for access for legitimate actors, independently screened according to objective criteria. This could exclude entities seeking data for purposes that violate human rights. However, they believe it should include NGOs and journalists with clear track-records. In general, the issue of access to whom, and to what extent, is not unique to Africa, although the question of dealing with data in many non-mainstream African languages may raise specific issues on the continent. Some informants recognise how UNESCO and others have proposed tiered degrees of access, such as with variations according to categories such as general citizens, journalists, accredited researchers, and official regulators as well as authorised law-enforcement. An African alliance for meaningful access to Intermediaries’ data-holdings would need to address the issues around authorised “clearing houses” to mediate supply-and-demand relationships.

There appears to be low awareness amongst informants that to the extent that (varied) API access exists (for certain jurisdictions) on the part of intermediaries such as Facebook, this is usually coupled to ethics clearance by universities. For Africa, this necessity would mean that stakeholders such as journalists and civil society groups conducting research need to partner with academics. Another option is partnerships with regulators like election management bodies, advertising standards authorities and child protection agencies in order to qualify for recognised access tiers – where these are, in theory, available.⁶ This situation points to co-operation among would-be

⁵ A list of further companies, and some of the issues involved from an external researcher point of view, can be found at Kamla Yadav and Alicia Wanless, “Social Listening Companies and Access to Sensitive Data,” Institute for Research on the Information Environment, May 26, 2022, <https://informationenvironment.org/wp-content/uploads/2022/09/PE2-Social-Listening-Companies-and-Access-to-Sensitive-Data.pdf>

⁶ To be noted, at the end of 2024, TikTok offered a degree of API access only to researchers in the US and the European Union. WhatsApp is owned by Meta, but only Facebook and Instagram are included in the parent company’s channels for researcher access. See Appendix C for more details.

stakeholders, and their engagement with policy-makers at national and Africa-wide level so as to bring that level of influence to bear on access.

4.3 Considering the sum of perceived barriers

When taken together, the barriers on both the supply and demand side for data access do not seem insurmountable. The benefits on all sides of having systems and mechanisms for openness and co-operation outweigh the disadvantages in the current situation. This includes making existing access channels meaningful in terms of data sets and categories (eg. in African languages). A new relationship could also help in everyone's interests as regards the quality of research findings about electronic communications in Africa more broadly, in contrast for example to the production of [overbroad claims based on miniscule data-sets](#).

Without change, however, the enduring problems will include ignorance of existing access channels, exclusion of African researchers from some platforms, diverse practices of informal and partial scraping (resulting in incomplete data sets), and exclusionary access via costly commercial brokers.

At the same time, some informants note that African stakeholders need to recognise that even in the event of securing access on the same basis as other jurisdictions, APIs typically have limited data sets available. This can restrict the ability of researchers to infer algorithmic recommender systems at work. Likewise, "likes" or "hashtag" data may reveal certain user behaviours and attitudes, but the platforms often do not share data on other less visible engagements – such as median time spent on the platform concerned, speed of scrolling, periodicity of pauses, etc. One informant noted that company data sets reflect the commercial interests of platforms (eg. in datafying mainly the signals that they can monetize), and may therefore have significant gaps from a different point of view.

There is a general agreement among informants that priority should be given to access to data to address pressing challenges related to elections on the continent. This entails data relating to disinformation, hate speech, digital authoritarianism, content distribution, and online gender-based violence.

Several informants for this research feel that an alliance for access to data in Africa should work to enhance a critical culture of datafication. It should also highlight the importance of access and use of already existing and available data on the continent by African researchers. Particular emphasis is put on the need for continental norms and standards that can help ensure protection of human rights, safety of data, affordability of datasets, and data flows.

A call for data access to combat online harms:

“In order to study any of the pathologies attributed to the transformation of the digital media ecosystem, social scientists need access to platform-controlled data as to ‘who’ saw or engaged with ‘what’ ‘when’. That is, scientists need to understand how and when certain people, and the population at large in different countries, interact with new media and what the consequences of those interactions are. Even when the platforms have promised to make available data for independent academic research, those promises have often gone unfulfilled.”

“Public authorities must compel major Internet platforms to provide independent parties with meaningful data about the impact social media has on democracy. In particular, platforms must:

- ❖ Share secure, privacy-protected data with certified academic institutions to examine issues such as:
 - auditing algorithms for bias towards extremism,
 - understanding the effect of social media on political polarisation and information consumption, and
 - disentangling the relationship between online hate speech and offline violence.”

Protecting Electoral Integrity in the Digital Age. The Report of the Kofi Annan Commission on Elections and Democracy in the Digital Age. January 2020

5. Global momentum and African prospects

On the data “supply” side some platforms have been narrowing existing access in several ways. One informant expressed frustration in regard to X for “the way things have shut down”. He argues that “public goods were piggybacking on the open API, but had to close down overnight”. Relative transparency was abruptly lost, and valuable services like “Bot or not” had been cut-off.

However, to comply with the European Union’s (EU) Digital Services Act, X in late November 2023, reinstated some access possibilities for researchers interested in online systemic risks to the EU. This is part of requirements for designated platforms under [European Union regulations](#), which will be mediated in each EU country by a national Digital Services Coordinator (DSC) of Establishment to be established as EU member states’ regulatory bodies.

There is also voluntarily opening up in selected areas (See Appendix C). This aligns with the large corporate members of the [Global Network Initiative](#) (including major international platforms and telcos) who have some voluntary momentum in assessing and addressing obstacles to data access through the [Action Coalition for Meaningful Transparency](#).

Back in June 2022, the major platforms committed to research access via an advisory “independent, third-party intermediary body” under the EU’s revised [Code of Practice on Disinformation](#). This Code includes provisions that companies will not prohibit, discourage or take “adversarial action” against public-interest good faith research into disinformation on their platforms.⁷ It also commits signatory platforms to provide fact-checkers with information and appropriate interfaces, to help them quantify the impact of fact-checked content, such as actions taken on the basis of that content, impressions, clicks, or interactions.

⁷ X is not a signatory of the EU’s Code of Practice on Disinformation (although, as Twitter, it was a signatory in 2016 of the EU’s [Code of Conduct on Countering Illegal Hate Speech Online](#)). X launched legal action in 2023 against the Center for Countering Digital Hate followed the NGO’s research that no action had been taken against most of a sample of Twitter Blue accounts which the Center had reported for “[tweeting hate](#).” Similarly, later in the year, X’s owner threatened a “[thermonuclear](#)” [legal case](#) against the US NGO Media Matters for reporting adverts adjacent to hate speech.

Big tech companies have further agreed to a [draft code of conduct](#) on researcher access to data, and the need for an independent mediatory body to adjudicate data access, as authorised by the EU's General Data Protection Regulation (GDPR). This draft code, adopted by platforms after extensive negotiations with the [European Digital Media Observatory](#), is accompanied by a model data-sharing agreement that would ensure privacy-compliance. The draft may ultimately be recognised by the EU.

While these issues are in process, there is an early attempt to see if Europe's policy has wider spin-off. This is by the Institute for Data, Democracy, and Politics at George Washington University, which has launched a [Platform Researcher Access Tools & The Brussels Effect Tracker](#). The [creators argue that](#) "Europe's actions may also help to undercut many of the platforms' arguments for not providing data to researchers."

Draft legislation in the US ([The Platform Accountability and Transparency Act](#)) may add to the EU impetus for openness. But generalised access via regulatory arrangements is not likely to happen in individual African countries in the short term. Yet these developments outside the continent could create a climate within Africa, and/or open up the possibility of partnerships between Africans and those abroad (including Africans in the diaspora) who are eligible to benefit from offshore developments.

On the other hand, informants are well aware that problems in connectivity and electricity constrain datafication in Africa, with cascading effects on what data sets exist and their reliability. Further, there is recognition that much collection and storage of data is controlled by international companies that operate beyond national jurisdictional authority. The legal and policy environment in many African countries is weak, and implementation is fraught.

However, an [expert meeting convened by RIA and partners](#) in Cape Town in November 2023, noted that the data access issue intersects with the implementation of several pertinent African normative instruments. The [report on this consultation](#) includes detail on further steps to be taken. Following up on this, communications were initiated with the African Rapporteur on Freedom of Expression and Access to Information, as well as members of the African Platform on Access to Information, the African Network of Information Commissioners and the others with a view to taking this issue further.

5.1 African stakeholders with interests/work in taking forward access to data

Relevant constituencies approached during this research included many of the following:

- ❖ Researchers and academic networks
- ❖ Science and research councils
- ❖ University alliances
- ❖ Civil society NGOs (eg. those monitoring hate and disinformation topics - and the impact of responses to these; health issues; crisis detection and response),
- ❖ Start-ups - eg. in the African AI-field
- ❖ African journalists and fact-checkers
- ❖ Regulators and their associations (elections, data protection, ATI, consumer protection, advertising)

5.2 Topics of interest:

Informants stress the importance of public interest purposes in accessing online data. Many pinpoint elections as a key public interest topic. More broadly, there is great interest in using data to track online disinformation and hate speech online, and intermediaries' responses to these kinds of expressions and related networks. Issues such as social divisions, prejudice and xenophobia are highlighted. In addition, there is interest in data relevant to online narratives about voter interest/apathy, corruption, health, education, safety, migrant-trafficking/smuggling, gender-equality climate and sustainability. Conflict prevention and peace-building are also raised as topics where access to online data could strengthen civil society's role.

Network analysis is important to many informants, especially when looking at orchestrated information operations and shares of voice. Data that reveals behaviours, not content in isolation, and flow paths of content, are seen as important. A number of informants want to research algorithmic recommender systems and amplification/de-amplification and shadow bans, which is not possible through scraping or commercial broker interfaces. One informant calls for transparency on algorithms which could be used to guide interventions that can break through filter bubbles.

There is also enthusiasm about online community mapping, and about identifying political party communities as clustered in echo chambers, in order to assess what they talk about, and being able to assess the ideology that underpins the narratives. This is seen as important for analysing levels of online polarisation. Another informant highlighted the importance of data that can help identify influencers and track their online impact.

5.3 Data ownership issues

There is interest in accessing metadata in regard to encrypted content, with strong awareness about protection of personal data and non-retention of data after a period. However, one informant believes that where there is data of small scale, anonymisation may be challenging.

In general, informants say that it is not about asking data holders to give out data recklessly, but about transparency regarding what data are collected, who gets access, how they are used and whether they are protected. Said one informant: "It is also about what Africans get in return for giving their data that others make money out of". A further point interviewees note is the fragmentation of the Internet into walled-gardens of data owned by diverse Intermediaries, making it difficult to track the bigger picture such as cross-platform trends in information operations.

One informant proposes that African regulators require open APIs for researchers and small enterprises, saying "it is not true that they [the platforms] are too big [to govern]". Another argues that outside of legal compulsion, ethically-speaking, intermediaries should make data they own available to not-for-profit entities such as those engaged in tolerance building. In this view, platforms should welcome the efforts of entities whose actions can help mitigate content risks on their services. One informant believes that the companies will not want to be profiled as unwilling or unhelpful data holders, especially on burning matters of public interest.

Related points from informants are to bring in "data subjects" to donate data for research, and to crowdsource data collection within countries. Concerns cover "extractive" collaborations from abroad with African researchers, and with the lack of research financing within Africa.

5.4 On mechanisms and systems for data access

There is little awareness among many informants of data trusts and stewardship issues as alternatives to bilateral access by individuals and platforms to privately-held big data. A few informants have had Crowdtangle access, but most not, and one says their organisation had applied without receiving a response. A number of informants believe that gaining generic API access is not at odds with partnerships on specific topic issues that provide wider access, as well as experimental research possibilities (as Meta allowed with its 2020 US research partners). Some informants who use APIs directly, or interfaces via data brokers, have built their own custom tools to extract and assess relevant data, which raises questions of possible duplication and fragmentation of effort and expense.

5.5 Capacity issues

Informants are well aware that access is one thing, and skills are another. Data analytics are needed - for example, to recognise whether one week's data is helpful, or if a year is needed. Says one informant: "Say you wanted to study the 2020 general elections in Tanzania, you would need large amounts of data. Not just 5000 posts".

According to informants, there is a challenge with regards coding and programming capacity, and access to data scientists very limited. These include skills of visualisation in the presentation of data-based findings. In some institutions, senior people may feel threatened by their lack of skill and understanding, and universities could experience a lack of qualified personnel to supervise student research in these areas.

The capacity to assess video as data is raised as an issue. This medium requires sophisticated machine learning models for thorough scrutiny and identification of harmful content, and for tracking the users disseminating it. One informant noted that purveyors of disinformation on platforms like Facebook and Twitter increasingly use images and videos to spread harmful content, with these formats being less easily detected due to similar technological constraints.

The expertise required for thorough data analysis comes at a high cost for independent not-for-profit organisations. This includes the skills of machine learning engineers, data scientists, and forensic researchers capable of extracting meaningful insights from large datasets.

Changes by platforms and changes/closures in freely available tools have meant that NGOs with resources have had to resort to coding their own tools. This helps to continue tracking a degree of public-facing data, but the informants concerned note that smaller operators cannot monitor to the same extent. One informant says: "Hardware is cheap; technical skills are not". Another says that newsrooms in particular do not have competence in collecting and analysing data at scale.

A further issue that informants raise is that stakeholders are not thinking enough about communicating their findings. "It is not just about expertise in developing infographics, but recognising the value of journalists working with coders on stories from the outset", says one.

5.6 Issues of focus for different stakeholder groups

The following are the various issues of interest for different stakeholder groups:

- ❖ Academics say they are interested in expanding their largely qualitative traditions into quantitative research activities in response to growing datafication and the rise of big data sets. An initiative to build research capacity in digital research methods has been launched.

Some private universities in Africa are making progress, especially in the field of economic analysis, but their public counterparts are seen to be lagging. Co-operation around a credit card company's data has been controversial. National research institutions and networks are important stakeholders. Some academics express that they need a community to call on, to share advice, skills and software tools. Shared data pools would also be of value to them. One informant sees a clear link between platform data access and Open Science and the related movement in Africa. Another says that open access to data and knowledge promotes and optimises scientific co-operation, and intersects with equitable access. Interoperability and reusability are important. Some academics are reluctant to open up their own data for wider benefit, and some informants argue that since lots of data are generated in Africa, two-way streets which include the platforms are needed to optimise public value.

- ❖ Parliamentarians see data access as key to help ensure accountability in digital governance in Africa. A governance framework to promote innovation depends on data access, covering problem-driven priorities, and also encompassing e-government and open data issues. Telcos, which are within most African jurisdictions, are important data holders to have in scope. There is a widely-shared view amongst informants that a pan-African approach is called for. One informant said that the network of African competition commissioners could be included.
- ❖ Election management bodies are interested in data access. The Association of African Election Authorities has adopted guidelines that call on platforms to provide “access to curated relevant data, including by means of application programme interfaces, to enable independent monitoring of content and networks that may harm election integrity”. They recognise the importance of partnerships to be able to analyse such data.
- ❖ Fact checkers and journalists see data access as essential if they are timeous in doing fact-checking and pre-bunking as rumours begin to get traction. This is key for early warning for the platforms themselves as well as societal stakeholders. Following through on the circulation of fact-checks is another area. Source checking via data access is vital as is identifying the actors behind the disinformation. Data availed by platforms should include the number of posts labelled/removed/downgraded broken down to geographic region and languages and the key topics and trends. The data that is most useful for fact-checkers to conduct network analysis includes:
 - Post data: This is the data on all public posts on the platforms, e.g. the dates, content, reach and interactions. This helps in the assessment of the authenticity of the content, coordinated amplification of content, etc.
 - User data: This includes all public profile information for users, which is instrumental in identifying inauthentic accounts and behaviours.
 - Advertisement data: This includes all public data available for advertisements to enable the identification of disinformation actors who use a platform's automated advertising services to amplify disinformation content.

Additionally, platforms can provide fact-checkers with access to data archives (with clear usage protocols) for better training in fact-checking and for understanding actor

behaviour. Insight is needed into steps taken against particularly egregious actors sanctioned for abusing platforms in order to gauge the impact of fact-checking.

Said one informant: “We would want to see how our fact checks are impacting and spreading. We’d like to delve into the data here to show reach. But that we don’t have access to.” This kind of information could assist in refining and sharpening fact-checking tools and how corrections are presented. Much fact-checking is prioritised based on what possible harm the item could present, meaning that engagement and virality alone are not the most important factors. However, while some data is given to fact-checkers about the extent that an individual item of questionable content is shared, the same is not available in terms of how the fact-checks circulate. When it comes to voice-note corrections on WhatsApp, it would be helpful to have metadata to track the flow.

- ❖ Regarding journalists, research by one informant’s organisation shows that despite the spread of viral and misleading information by many political actors and other influencers, journalists seldom come across such content because they have a weak understanding of the phenomenon. Traditional newsrooms have low adoption of opportunities to do data analysis due to fears of working with numbers which is a disincentive. The data journalism desk at a major news outlet has been closed down. On the other hand, according to one informant, “it has become necessary that the value journalists add to public conversation has to be obvious and distinct, and being data competent is one way to do this”.

Journalists express further that it is hard to do a story off tech-data alone, and that offline sourcing is also needed. However, according to Code4Africa, in some 23 countries, journalists are using digital tools to analyse data, including content “laundering”. Data forensic journalism using OSINT and scraping is taking place. For data-journalists, official statistics and large data sets from online African news media operations are currently the main data source. “Sentinels” are used to flag problematic content for news reporting upon, for example as this content circulates on WhatsApp. A number of projects are supported by Google with financial contributions, but not data. Code4Africa has 18 university partners, which help provide peer review and validation. Africa Data Hub (ADH) is a data platform for storytelling around health and development based on verified up-to-date data, and is combined with tools, training, and mentoring support.

Data about attacks on journalists, especially women, and women in politics is further area of interest to media and NGO actors. They point out that it is hard to get context and to track build-up, however, without access to data (and/or meta-data).

- ❖ Bulk access to data in African indigenous languages is very relevant to start-ups working on large language models that could be of use in content moderation, customer service and e-governance applications. In the perception of one informant, “big tech” are not using their data in these areas, opening the door therefore for partnerships based on innovation outside the main players. However, current limits restrict even experimentation, as well as teaching/training purposes. A data scientist in this space says that what becomes difficult is the demand by data holders to explain what is sought for collection and why, even before a researcher really even knows the range of data that would be productive to harvest in regard to developing a particular application.

- ❖ Civic tech personnel advocate for a wider approach that includes Africans collecting their own purposive data using offline tools, and encompassing people with low integration with platform datafication operations. In the words of one key informant, “we should collect what they (the platforms) don’t have”.
- ❖ UN actors working in Africa point to the value of platform data to their work in combatting disinformation, noting also that their data needs and related “asks” evolve over time.

6. Advantages for improved access

On the companies’ side:

- ❖ Addressing criticism of global imbalances in treatment of Global South, platforms can invest in access to data for Africa, such as in terms of API access, as well as capacity building.
- ❖ This modality of transparency will set an example for governmental transparency, and for partnerships around states availing official data.
- ❖ By providing carefully curated access to data, companies will reduce pressure on them to provide “backdoors” to authorities that are open to abuse.
- ❖ Platforms can benefit from African researcher findings and even data by improving openness and relationships with stakeholders.

On African users’ side:

- ❖ A balance can be struck between personal data protection and data sharing, with appropriate attention given to the latter side of the coin.
- ❖ Approved diverse sectoral entities, sometimes in concert, can process data for the purpose of developing public-interest insights relevant to African democracy, development and scientific advancement.
- ❖ A cadre of African users can take their place alongside others elsewhere in the world who already have, or are soon expected to have, improved access to intermediaries’ data.
- ❖ African researchers are interested not only in parity with counterparts elsewhere, but in pursuing active partnerships as well.

7. Conclusion and recommendations:

Improving the current access situation could be achieved through building an alliance of interested parties to engage with each other and with the range of data-holders relevant to Africa. Considered individually, the diverse actors on the demand side are fragmented and constrained in terms of approaching Intermediaries, not to mention having limited impact on national and continental policy about platform data access.

Experience elsewhere shows that joint approaches can yield momentum. This is the case with the European Digital Media Observatory (EDMO), which generated a code of conduct for researcher access to data under the EU’s General Data Protection Regulation. The USA has seen the formation [of an academic coalition](#) to advocate for the issue. In the same country, the Centre for Democratic Technology [brought together a group of US researchers to brainstorm ideal research agendas](#). The [United Nations Inter-Agency Group on Disinformation](#) has developed generic data asks from

platform companies. Under the auspices of 45 countries, the [Observatory on Information and Democracy](#) is working to advance a worldwide call for access to Intermediary data.

This research reflects the perspectives of underrepresented groups in global digital transparency conversations. For Africa, transparency and data access go beyond the discussions about how these matters can help enhance accountability and contribute to informed regulation: there is also, on the continent, huge democratic and developmental value of such access.

At the same time, it is also underlined that supply side access is just one part of the equation: attention is also needed to wider awareness-raising, data science empowerment, networking and formulating a shared approach to the issues.

In a follow-up stage to this research, it is suggested that:

- ❖ Ongoing networking and training is needed to build capacity to meaningfully identify and process particular data sets, both on existing access arrangements or future, deeper ones.
- ❖ Roundtable dialogues such as about data and elections can brainstorm specific modularised data asks⁸, as well as aggregate individual interests, develop joint initiatives, and propose mechanisms for vetting and quality control.
- ❖ These activities should be integrated with discussions of representatives of the platforms and extend to other private and public data holders as well.
- ❖ Identify possible pilot exercises involving the key African stakeholders (public sector and civil society) and local representatives of technology companies.
- ❖ March 2-8, 2024, is Open Data Day, where previous years have seen almost 30 celebrations in Africa. This is an occasion to advocate for data access from platforms as important to African democracy and development.
- ❖ Possible engagement with the sixth African Science Technology and Innovation Forum, expected to be convened in 2024 by the UN Economic Commission for Africa.
- ❖ Possible engagement with the Council for the Development of Social Science Research in Africa (Codesria) about events to engage academics and national research institutes further.
- ❖ Sharing costs of software development and ensuring visibility imperatives should result in collaboration between researchers, and between them and journalists.
- ❖ The insights here should be integrated into the ongoing discussions with the African Union, and sub-regional African bodies, so as to develop practical alignments with the implementation dimensions of the Data Policy Framework.
- ❖ The pertinent African Union instruments should be linked together into implementation guidelines applicable to different stakeholders.
- ❖ This document is brought to the attention of international actors outside the African continent, who may see opportunities to support relevant data access and capacitation within their parts of the world, in regard to African-relevant data and African actors.

⁸ See for example: UNESCO. 2023. [Data sharing to foster information as a public good. The case of media viability and safety of journalists in the digital ecosystem](#)

Appendix A:

Key informants interviewed:

1. Kyle Findlay, Murmur Intelligence
2. Aldu Cornelissen, Murmur Intelligence
3. Laura Grant, Outlier
4. Alistair Otto, Outlier
5. Joshua Olufemi, Dataphyte
6. Michael Salzwedel, Internews
7. Herman Wasserman, University of Stellenbosch
8. Francis Nymanjoh, University of Cape Town
9. David Cooper, University of Cape Town
10. Camaren Peter, Centre for Analytics and Behavioural Change
11. Caroline Anipah, Dubawa
12. Chris Roper, Code4Africa
13. Andrew Rens, RIA
14. Scott Timcke, RIA
15. Doreen Wainainah, PesaCheck
16. Noko Makgato, AfricaCheck
17. Heide Hackmann, University of Pretoria, Chairperson, The African Open Science Platform.
18. Nuzulack Dausen, Nukta
19. Alette Schoon, Rhodes University
20. William Bird, Media Monitoring Africa
21. Neema Lungungira, African Network of Parliamentarians on Internet Governance
22. Patricia Adusei Poku, Network of African Data Protection Authorities
23. Wairagala Wakabi, Cipesa
24. Godwin Murunga, Codesria
25. Churchill Otieno, African Media Convention
26. Gabriella Razzano, OpenUp
27. Vukosi Marivate, University of Pretoria,
28. Daya Reddy, University of Cape Town.
29. Phumzile van Damme, consultant
30. Gilbert Sendugwa, Africa Freedom of Information Centre

31. Lilian Nalwoga, Cipesa
32. Steve Vosloo, UNICEF
33. Dorothy Gordon, Information for All Programme, UNESCO
34. Ambassador Bitange Ndemo, Kenyan government
35. Philip Ayazika, Pollicy

Appendix B:

Consultations and discussions:

In Abuja, [side event at the Africa IGF](#), 19 September 2023, based on [this discussion document and questions](#).

In Dar es Salaam, side event at the FifAfrica, based also on the above discussion document and questions. 15 participants from digital rights organisations and individual digital rights activists attended the meeting.

In Cape Town, some 50 participants attended [workshops on 16 and 17 November](#), centred on elections and data. Convened within the wider framework of this initiative, they brought into common dialogue different stakeholders (including individual key informants). The events [produced strengthened awareness and understanding, as well as outcomes calling for further specific actions](#).

Appendix C

Twitter/X:

The company's formerly open research API was scrapped in 2023, and replaced with an [expensive payment regime](#). Data access for African users [has suffered directly](#) – as well as indirectly (eg. [such as via use of bot monitoring apps](#)). However, one entity interviewed for this research said it had paid the company a relatively low sum for specific access (\$50 for a month's limited access). To comply with the EU's Digital Services Act, X in December 2023 offered [a degree of limited access](#).

Meta:

A very small number of Africans have had access to a limited degree of data (eg. [engagement, but not reach](#)) from Facebook and Instagram via the Crowdtangle interface, but in 2023 this facility stopped giving access to new users. Existing African users report continuing disruptions in the service.

Meta's [Facebook Open Research & Transparency](#) (FORT) for approved academic researchers plus ["others considered on a case by case basis"](#), and offers a degree of free compute capacity. The company says that this access concentrates on elections and democracy as a research area. Its [researcher platform](#) provides a Researcher API allowing access to ["near real-time data as well as billions of historical data points"](#). It covers "data from all public Facebook Pages, Groups, Posts and Events", incorporating URL shares, but the website also notes that this is [limited to the US and select EU countries](#).⁹ In September 2023, the [website said it](#) was "not accepting new partner applications at this time".

Earlier, in August 2023, the company [announced](#) a ["Meta Content Library and API"](#). This facility is said by the company to include data [in 180+ languages](#) from public posts, pages, groups, and events on Facebook, and for Instagram, public posts and data from creator and business accounts. Data from the Library can be searched, explored, and processed via a [web-based controlled-access interface](#) which does not allow downloading of data (although research outputs can be exported). This limitation on export also appears to be the case with data for deeper analysis that is possible through the company's Researcher programme, which is done by approved access to the Content Library's API. The Library offers search across over 100 data fields from Instagram accounts and posts and Facebook Pages, posts, groups and events, with filtering for language, view count, media type, and content producer.

Although one African country ([Togo](#)) is amongst some countries' data that are excluded for "legal and compliance" reasons, "global researchers" are eligible to apply if they are affiliated to [an academic institution or a qualified research institute](#), and applications are reviewed by the [Inter-university Consortium for Political and Social Research \(ICPSR\)](#) at the University of Michigan, US¹⁰. The data-sets entailed in the 17 researcher example cited earlier are being archived by the

⁹ Approved researchers have access to a US' ["Civic Engagement Data Set"](#) that includes aggregated demographic statistics of user engagement (Likes, Comments and other actions associated with content) with civic and voting features (Voter Turnout Reminder tool, Share That You Voted tool and Voter Registration tool) on Facebook starting on February 11, 2020 and up to and including December 31, 2020.

¹⁰ An amount of \$50 000 Amazon Web Services credits have been donated to the University of Michigan's Institute for Social Research to offset computing costs for their Social Media Archive.

same Consortium, in which 2 Ghanian and 6 South African entities are among the 750 mainly north American members.¹¹

On [advertising data](#), Meta [also says](#) it offers “granular ad-level targeting information for all social issues, electoral and political ads run across Meta's technologies since August 2020 in 120+ countries”. This dataset includes “ad targeting criteria, such as age, gender, location, Custom Audience and Lookalike Audience inclusion or exclusion, and other detailed targeting selections (demographics, behaviors, or interests).” The company’s [Ad Library API](#) enables a deeper analysis of ads about social issues, elections or politics, as well as all other ads that deliver to the EU. The access to data (outside the EU) [includes](#) targeting data for social, electoral and political ads (such as location, age, gender, detailed targeting, language, custom audiences, and lookalike audiences); impression range, and the reach percentage of people who saw an ad, by age and gender, and information about the locations where the ad reached; as well as spend.

While Meta’s encrypted messaging system WhatsApp is enormously present in Africa, the service appears so far to be loathe to give any access to areas of its massive meta-data holdings (this applies outside Africa as well). Meta uses this meta-data to add to its advert profiling for Facebook and Instagram. This data-set could however be important for crisis detection, management and response. While Whatsapp can be [a useful tool for research](#), and a [means of data collection](#), this is a different matter as compared to accessing (meta)data about its use and the flow of particular messaging. The WhatsApp Business Platform API allows paying businesses to “connect thousands of agents and bots to interact with customers programmatically and manually”, which signals the utility of certain metadata for narrow commercial users. Again, the significance can be inferred (and elaborated) from the affordance whereby this business API can enable a business [to retrieve all conversation](#) with the specific phone number/s they are using, by time period and country.

Meta provides special access to US academics, as evident in [the company’s provision of access for 17 US academics](#) and two research surveys through [a research interface](#) during the 2020 US elections.¹² This raises the question of: if there, why not also in African countries?

Google:

While Google trends and some ad-tech data is available, this is without more granular offerings. Google-owned YouTube has [a researcher program](#), which - like other companies - effectively gives the platforms sole power to screen researchers and research projects, and is limited to public data. The data-sets available include ratings for a given video, sharings, and marking a video as a favourite, among others. To be approved, researchers must be linked to an accredited university, “have a clear research goal, and intend to publish their findings”. Topics have to be [approved by YouTube](#). Academic researchers in [some 60 countries](#), including Ghana, Kenya, Senegal and South Africa, are eligible to apply for access to the API. [The YouTube researcher terms](#) also acknowledge eligibility of “any government or other institution required by law or regulation to have access to Program Data”. Specific information about African uptake of this access by academics or others could not be found for this research.

¹¹ The company says the contracts it signed were the same as “other independent researchers who use our data, which is publicly posted on [Social Science One’s web site](#)”.

¹² Four studies published in Open Access in Nature and Science journals in mid-2023 covered: How do social media feed algorithms affect attitudes and behaviour in an election campaign?; Asymmetric Ideological Segregation in Exposure to Political News on Facebook; Reshares on social media amplify political news but do not detectably affect beliefs or opinions; Like-minded sources on Facebook are prevalent but not polarising.

TikTok:

Some African researchers using TikTok data appear to be generally scraping it in a piecemeal fashion. The company in 2023 was offering some [API access](#) to researchers at US- and European universities, such as numbers of video followers and likes. It approves proposals and then researchers have to abide by quite restrictive [terms of service](#). The data availed is [assessed by some as limited](#). Problems with the TikTok API [include constant changes](#). It is not evident if/when any access will be extended to Africans.