WHITEPAPER

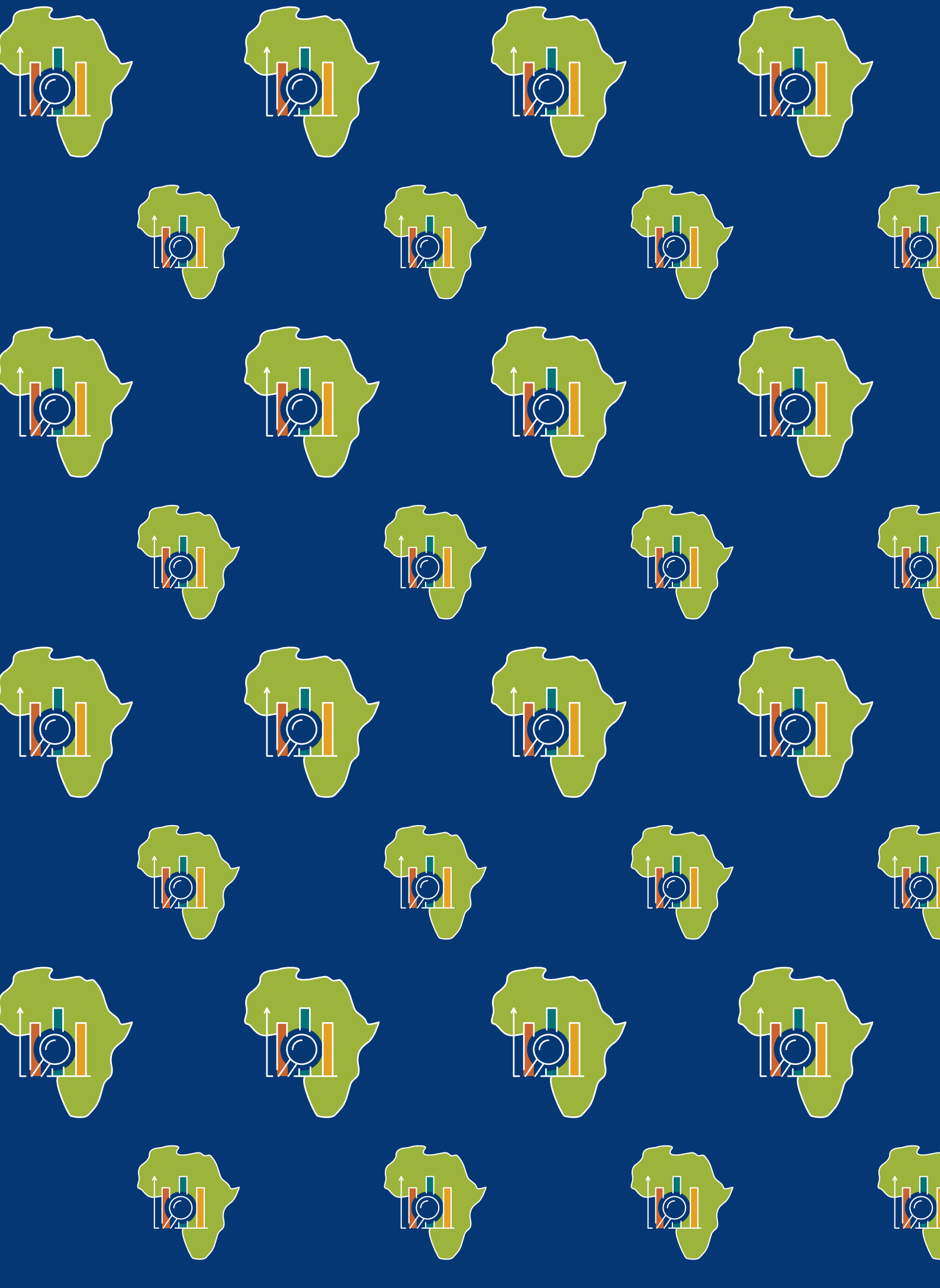
FINDING QUALITY DATA IN AFRICA

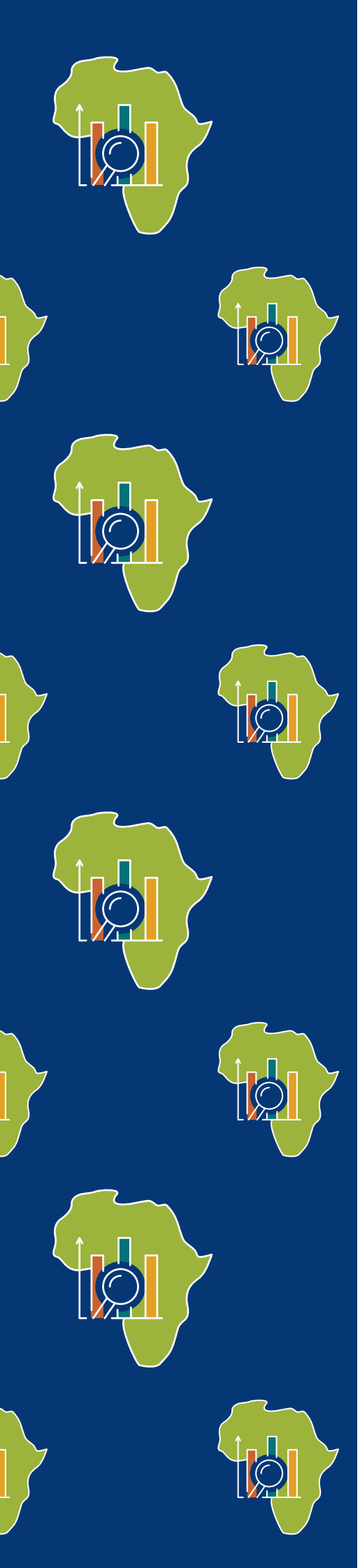
LESSONS FROM THE FIELD

BY HELENA BARNARD, JOHN M. LUIZ, FANG LEE COOKE

Gordon Institute of Business Science
University of Pretoria

2025





GIBS www.gibs.co.za

Contents

GIBS **AUTHORS**

HELENA BARNARD
Gordon Institute of Business Science, University of Pretoria, South Africa
Email: barnardh@gibs.co.za
https://orcid.org/0000-0002-5690-5023

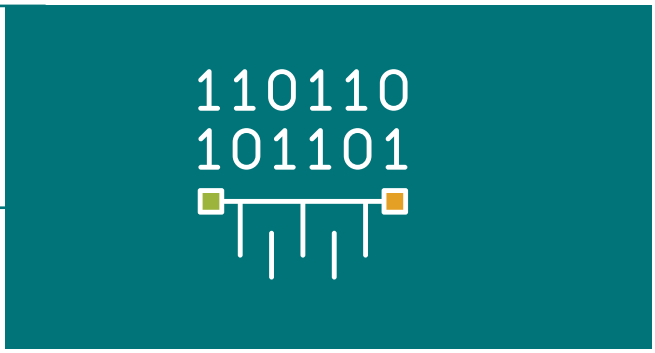
JOHN M. LUIZ
University of Sussex Business School, UK
Email: J.M.Luiz@sussex.ac.uk
https://orcid.org/0000-0003-3806-4424

FANG LEE COOKE
Monash University, Australia
Email: Fang.Cooke@monash.edu
https://orcid.org/0000-0003-0337-6591

Abstract

As data becomes more important for decision-makers, it also becomes increasingly critical to ensure data quality. However, in Africa, getting data can be hard – not to mention getting high-quality data. Quantitative data is often outdated or incomplete, and even if data can be found, it is too limited for the intended purpose. Gathering qualitative data is often complicated by local cultural norms and the lack of education of respondents. Experienced scholars from both inside and outside the continent reflect on some of the challenges they experienced seeking to gather data in Africa and share their strategies to help others gather better data.

Introduction



"Data" has become almost a buzzword, as many managers try to make more evidence-based decisions. Good quality data is the starting point for many strategic decisions. Yet, in Africa, data-driven decision-makers often struggle to obtain robust data, whether they are looking for quantitative data that can be statistically analysed or whether they seek to gather and understand qualitative data. Data quality in Africa cannot be assumed, but this does not mean that it is impossible to get solid data for decisions. There are strategies that can be used to improve the quality of data.

Emerging markets are virtually, per definition, undergoing rapid and far-reaching institutional changes (Marquis & Raynard, 2015). Although the growth potential of emerging markets is often emphasised, these rapid changes are also often accompanied by political instability and institutional weaknesses. Add into the mix the reality of economic constraints and changing economic priorities, and it is hardly surprising that the exacting but unglamorous task of systematic data gathering often falls by the wayside.

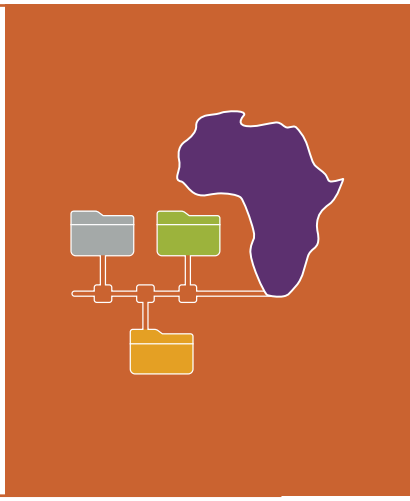
It is precisely because of these changes that what was valid a few years ago may no longer apply. To support decision-making, empirical evidence should be timely and able to map changes over time.

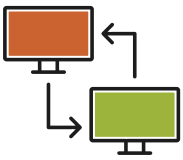
Furthermore, considering the context is changing so fast, it is challenging to obtain longitudinal evidence – government departments merge and are renamed, and firms migrate (Barnard, 2014) or are acquired. This changeability places particular demands on people trying to find quality data in emerging markets like Africa.

In this white paper, two scholars share their experiences and their insights. Prof. John Luiz is South African born and raised and has long studied institutions in Africa.[1] He comments about his experiences gathering quantitative data. Prof. Fang Cooke was born and educated in China,[2] then further educated and worked in the United Kingdom for 16 years before emigrating to Australia. She seeks to understand people and their workplaces in different country contexts – for example, diversity and employment relationships in China and India. Moreover, Prof. Cooked has studied the employees of Chinese firms in Africa, and reflects on her experiences gathering qualitative evidence during those projects. Both scholars share the strategies they have developed to improve their data collection.

[1] John Da Silva Luiz Profile | University of Sussex
[2] Fang Lee Cooke - Monash University

Seeking to construct robust quantitative datasets in African countries





Data availability

Early on, Prof. Luiz learnt how challenging it could be to seek primary data in institutionally turbulent environments. In 1996, he was trying to visit the Central Bank of the then Zaire for data collection. When he phoned, he was told not to bother because data was the least of their concerns; they were preparing to shut down as the rebel armies were approaching. A few weeks later, the Mobutu regime fell.

Even in South Africa, an upper-middle-income country, the situation was difficult. Prof. Luiz served for six years on the Statistics South Africa Council and witnessed first-hand the enormous resource pressure the organisation experienced to just do the basics (current gross domestic product [GDP], inflation data, etc.). It was clear that robust firm-level data was not a priority.

International databases like Orbis are often seen as credible sources of information. However, the limitations of Orbis become apparent when cross-referenced against other data sources like the Johannesburg Stock Exchange. Other scholars have also raised questions about such databases when it comes to less developed contexts (McGuire et al., 2016). Similar concerns can be noted about the World Bank Enterprise Surveys, a frequently used source for firm-level data in African studies. These surveys are sporadic and thus often outdated – the most recent survey conducted in Nigeria was in 2014, and before then in 2007. Moreover, the nature of these surveys can result in sampling bias, as they often do not cover all regions or sectors within a country (Gelb et al., 2007).

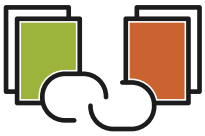
Even macroeconomic data is problematic

The problems associated with the quality of statistical data being produced by the national statistical bodies in Africa are well-known. For example, Jerven (2013, p. xiii) wrote about how we are "misled" by African development statistics and that decisions are being made based on "dubious" statistics because the data is either inaccurate or non-existent. He cited the examples of Ghana and Nigeria, which revised their GDP estimates upwards by over 60% and 89% in 2010 and 2014 respectively. This resulted in both countries being reclassified from low to lower-middle-income countries. These two countries are amongst the 10 largest economies on the continent and data quality is worse in many of the smaller and less developed African countries.

In 2011, a World Bank chief economist for Africa (Devarajan, 2011) wrote a piece titled "Africa's statistical tragedy", stating that much data used by economists needs to "be taken with a grain of salt" and that, for many African countries, "we

simply don't know". Documenting some of the causes for these data problems, Devarajan (2011) cited: "weak capacity in countries to collect, manage and disseminate data; inadequate funding; diffuse responsibilities; and fragmentation".

Additionally, Jerven (2013) indicated that there is a lack of equivalence of data and methods in national accounting practices across countries on the continent, making it difficult to do comparisons across African countries.



Challenges of collecting longitudinal data

Since it is virtually definitional that emerging markets are undergoing rapid change, longitudinal data is essential for scholars who wish to understand those countries and businesses seeking to make projections. Prof. Luiz and some of his colleagues decided to construct their own dataset to analyse South Africa's economic and institutional development before, during, and after apartheid – an ambitious century-long time period. Rather than taking three years as expected, it ended up taking 12 years – most of it dedicated to the painstaking construction of a dataset through archival work. Under apartheid, data was often remarkably detailed, but scattered. For example, education records were spread across over 20 different departments that were racially and ethnically separated. As the country changed, the departments (and hence the data) seemed to disappear. It required astute investigation (and then back-breaking work) to find the data when reappearing in another form.

Precariousness of data

Prof. Luiz and his colleagues ended up with a comprehensive longitudinal dataset from which they published over a dozen papers (see Fedderke et al., 2001, 2006, 2008; Fedderke & Luiz, 2002, 2007). They asked the national statistics body of the country to take over the data, but Statistics South Africa did not have the capacity to integrate this historical data. These were the days before cloud storage; as the team changed, computers and programs changed, and the data gradually disappeared.

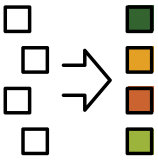
This was not an isolated occurrence. Since Africa – like most emerging markets – has such limited resources, data is often stored under precarious conditions. In 2021, a devastating fire at the Jagger Library at the University of Cape Town destroyed irreplaceable material stored in its Africa Special Collections. As academics, data-savvy decision-makers and engaged citizens, we all need to do more to protect data.

Challenges	Manifestations
Firm-level data is often unavailable or unreliable	• Often collected by national statistics agencies that lack resources • Surveys undertaken by international agencies, such as the World Bank, are sporadic and often unaligned to IB questions • Data coverage from major databases is often unreliable
Even macroeconomic data can be problematic	• Data is sometimes subject to major revisions years after release • Large gaps are evident in basic macro data, such as inequality and education data
Worldwide comparative data often poorly covers African countries	• For example, Hofstede's cultural dimensions data excludes many African countries • Some institutional indicator datasets are incomplete
Precariousness of data	• Data backups are not always in place and data is insufficiently protected

Table 1: Issues with quantitative data

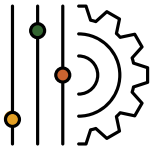
Recommendations for more robust quantitative data in under-researched contexts

Doing quantitative research in a context like Africa often leaves people with little choice but to use the flawed available data. However, they need to do so with caution. Alternatively, they can seek to develop and validate alternative sources of evidence.



When firm-level data is unavailable or unreliable

Given the lack of alternatives, World Bank Enterprise Surveys are an important resource. However, they tend to focus on access to finance and infrastructure and issues related to the regulatory environment, and should not be stretched beyond what they can contribute. Being mindful of the caveats of using such data and communicating this awareness is key to making this evidence useful.



When macroeconomic data is problematic

Scholars of Africa and other under-researched contexts may need to develop innovative data sources. Barnard (2020) mentioned flight data as evidence of global connectedness. The late Hamilton Ratshefola said he knew the estimates of earnings in the Democratic Republic of Congo were wrong when he saw the activity in the markets of Kinshasa. As he sought to launch his business there, he instead used the country-level earnings of MTN to estimate market size.

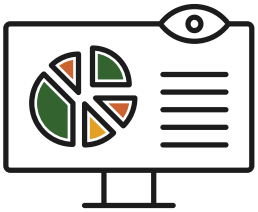
Similarly, companies and scholars may need to develop innovative proxies of the relevant indicators. Delays at border crossings are not only a signal of inadequate administrative capacity. Companies can note the length of delays and the types of trucks waiting for clearance to develop a better understanding of market activity in the receiving country. In agrarian countries especially, paying attention to harvesting cycles and local festivals can help businesses anticipate seasonal buying patterns. Informal traders are often invaluable sources of information about business sentiment.

When worldwide comparative data poorly covers African countries

"Digital trace data", the qualitative and quantitative evidence left behind by users of online platforms, provides the benefit of cross-national coverage. Even though bias effects have been noted (Liang et al., 2024) and emerging markets lag behind their more developed counterparts in the development of online content (Ojanperä et al., 2017), platforms capture usage by vast numbers of people across the globe. In many cases, the digital traces of business activities can be used to assess cross-national patterns, including in under-researched contexts. Recent examples include work on ride-sharing (Basukie et al., 2020) and online job-seeking platforms (Lehdonvirta et al., 2019), but many more sources can be used.

Considerable digital trace data can be obtained from private service providers like MPesa, MasterCard, and DHL. Scholars or businesses can demonstrate to these providers that it will be in their interest to share their data and they may well prove amenable to sharing the evidence. Various smaller-scale projects have already been undertaken (Saville, 2017), but much more can be done. In the African context, social media platforms like WhatsApp and Facebook groups provide a valuable and direct link with customers. The insights emanating from user-generated sources like generalist platforms (e.g., Twitter/X) and specialist sites like GitHub or LinkedIn also remain under-explored.

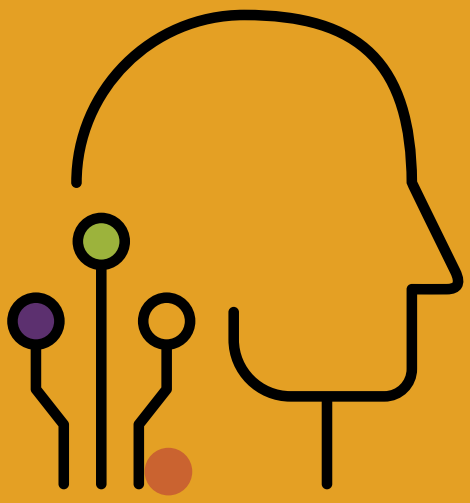
We live in the era of artificial intelligence (AI) and businesses can also leverage AI to support their data efforts. A central challenge is that Africans have been marginal to the development of AI and it does not necessarily reflect the African world (Marivate, 2024). Nonetheless, AI-driven analytics can be very useful in processing and interpreting unstructured datasets more efficiently, uncovering patterns and insights that might be missed through traditional methods.



Dealing with precarious data

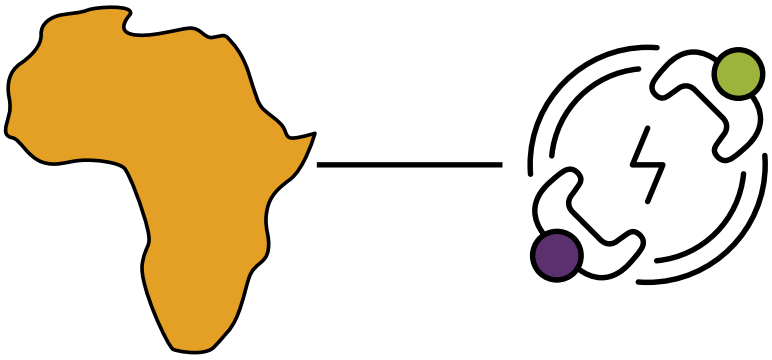
Considering data is scarce, it should be treated as a public good and shared on open platforms. However, data repositories are rarely user-friendly. A welcome development is around data journals, journals that publish papers describing the details of the collection, processing, software, file formats, etc. of a dataset (rather than its analysis). In the field of business and management, *Data in Brief* (https://www.sciencedirect.com/journal/data-in-brief) is a potential outlet for those who have taken the time to gather data – and a resource for others.

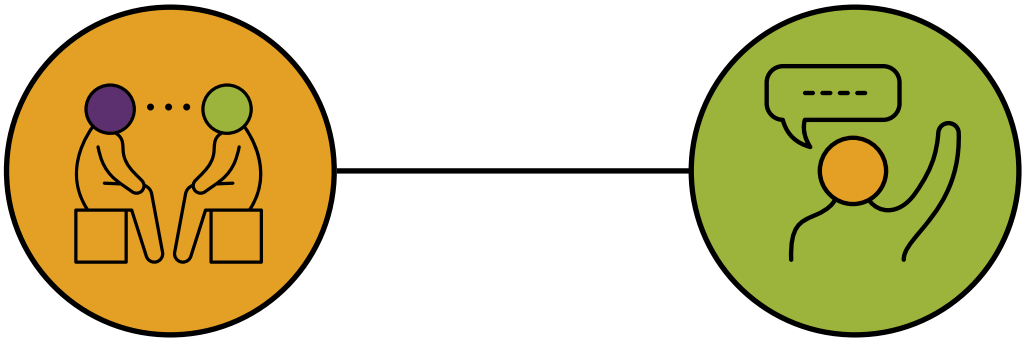
Seeking to conduct rich qualitative interviews in Africa



Qualitative data is often seen as an alternative in under-researched contexts, promising a rich understanding of phenomena (Creswell, 2009). The "messiness" and ethical challenges of qualitative research are generally acknowledged (Nair, 2021). Across various countries and employee groups, managers and workers often avoid disclosing company problems to protect its reputation or out of fear of repercussions. This reluctance complicates data collection and reflects the social desirability bias noted in research settings globally (Bradburn & Sudman, 1979).

Furthermore, it seems that these challenges are more serious and difficult to overcome when interviewing workers at the grassroots level in under-researched contexts. With African countries ranked very low on human capital indicators, interviewees' ability to understand interview questions and articulate their thoughts on the phenomenon of research cannot be assumed. Likewise, in countries with authoritarian regimes, participants may fear contradicting authorities. This may negatively affect their willingness to disclose information and share their thoughts. Given the high-power distance in many African (and other emerging market) countries, there is often a cultural norm of politeness. Researchers may find themselves politely stonewalled. Prof. Cooke experienced these challenges when researching the business motivations of the Chinese migrant entrepreneurs of small businesses in South Africa and the employment terms and conditions of their workers, largely undocumented migrants from other African countries (Cooke et al., 2022; Wood & Cooke, 2023).





Limited ability of interviewees to engage with the research topic

The workers had a relatively low education level and given that English was not their mother tongue, they struggled to express themselves. The researcher, in turn, struggled to understand unfamiliar accents and linguistic characteristics. Respondents seemed to not previously have considered the research issues and struggled to organise their thoughts. This problem persisted for some less-educated migrant entrepreneurs, even when Prof. Cooke switched to Chinese for the interviews.

Unwillingness of interviewees to disclose thoughts

Even when language was not a barrier, some interviewees were reluctant to disclose accurate information or share detailed thoughts. Despite careful explanations of the study's aims and assurances of confidentiality, workers were unwilling to disclose their wages and feared reporting poor management practices. Instead, they sometimes pretended not to understand questions or gave random answers.

This is not limited to Africa, as Prof. Cooke realised when she conducted a case study with a subsidiary of a multinational garment manufacturing corporation in Sri Lanka. The company had won a national award for its corporate social responsibility and interviewees were fairly relaxed about sharing the positive aspects of the company. However, they remained reluctant to disclose what challenges and barriers the company encountered when implementing new management techniques (Cooke, 2018). They just politely said a few things without giving a proper answer.

Barriers	Manifestations
Limited ability of interviewees to engage with the research topic	• Unable to understand the topic despite careful explanation • Unable to organise and articulate their thoughts • Becoming uncomfortable or embarrassed
Unwillingness of interviewees to share their thoughts	• Giving very brief or vague answers • Not answering the questions directly by giving random answers and misunderstanding the question deliberately • Become annoyed when asked the question repeatedly

Table 2: Barriers to collecting rich qualitative data

Recommendations for more robust qualitative data in under-researched contexts

The challenges experienced in doing interviews with respondents in Africa showcase the heavy reliance of the field on formal interviews and the spoken word as a source of data. Our recommendations suggest that a different approach may be useful.

Limited ability of interviewees to engage with the research topic

It matters whether interviewees cannot or will not respond to questions – that is, whether the challenge is one of ability or of motivation. Researchers should seek to understand which one it is. If interviewees are willing to share their views but struggle to do so, the researcher can separate the first engagement (where the questions are asked) from a second engagement (where interviewees answer). The gap can provide time for interviewees to think about and develop their answers.

Unwillingness of interviewees to share their thoughts

Voluntary participation is a cornerstone of ethical research, but sometimes interviewees do not want views attributed to them. This can be especially severe in oppressive regimes. A researcher can then calibrate information with that reported by others (e.g., the shop owner, workers' manager or facility manager of the retail outlet). Visual clues like job advertisements for similar jobs or online billboards can also be used. Where appropriate and reliable, secondary statistical data could be used to supplement the primary data. Ethnographic approaches where researchers spend more time with respondents can often help parties develop a more trusting and mutually enabling relationship.

General recommendations for more robust data from under-researched contexts

Most of the data-related challenges experienced in Africa are not fundamentally different to those found elsewhere, but they are often more severe. They provide a useful starting point for thinking about robust data generally.

To get in-depth, well-triangulated, and high-quality data, it is critical to understand the context of the work, the background of the phenomenon investigated, and the assumptions that are made in the process. Moreover, this needs to be communicated so that others can see and interrogate those building blocks of the data-gathering process. Expect to be challenged when not using apparently gold-standard databases and make sure to explain why those databases do not function as well in their context. For scholars, this typically means quite a detailed methodology section. For managers, it requires getting buy-in that the proxies are likely to provide more useful information than other measures.

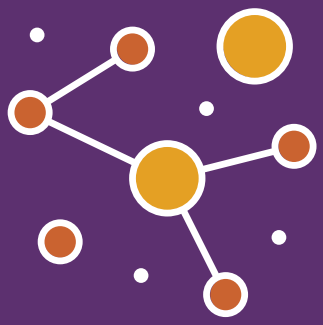
The precariousness – of livelihoods, firm profitability, institutional transformation and more – that so often characterises Africa, often also characterises its data. However, there is no reason to not insist on the most advanced analytical strategies. While companies and academics may not have much control over the type and quality data they can obtain, they can and should analyse the data in the most robust way possible.

As people who want to better understand and do better business in Africa, we should be doing more to develop, store, and share data. Having others use the same data not only represents the improved use of a scarce resource, but also is another quality check. We should be putting pressure on development agencies to support statistical capacity development and to ensure that data is openly accessible. This is especially true for macro-level quantitative data, where gathering, storing, and sharing data can be done in a more standardised and transparent way.

There is also value in firms collaborating to obtain better baseline data. Bearing in mind that firms need to deal with competition and concerns about anti-competitive behaviour, these benefits may be realised especially by firms that fulfil different roles in a business ecosystem. For example, a transportation company, financial institution, and selected manufacturers and retailers can work together to pool insights about markets that are data-poor. Such networks are known to be useful in supporting internationalisation (Chipp et al., 2019) and are especially useful when internationalising. By consolidating data across different industries, a 360-degree perspective can be developed.

For qualitative data, that may not be as easy. Where it exists, rich qualitative data is less easy to share because of the cost of desensitisation and the difficulties in interpreting the data without understanding the field setting – for which being involved in the fieldwork is often a prerequisite. What can be shared is best practices, namely: how to plan a project, how to use intermediaries, how to approach interviewees, and the like. This will help everyone have more productive fieldwork experiences.

Conclusion



If we accept that better data is the foundation of more informed decisions, then getting quality data for business in Africa is a business imperative. Sometimes, businesses just "shoot from the hip" and express surprise when results are disappointing. Other times, businesses unthinkingly use the datasets that are the norm in the rest of the world – even though those sources do not do a good job of covering Africa. It is not easy to obtain quality data about business in Africa, but we ignore the imperative of doing so at our peril. We hope that our experiences and insights inspire others to undertake this challenging but exciting task.

References

Barnard, H. (2014). Migrating EMNCs and the theory of the multinational. In A. Cuervo-Cazurra, & R. Ramamurti (Eds.), *Understanding multinationals from emerging markets* (pp. 195–223). Cambridge University Press.

Barnard, H. (2020). The Africa we want and the Africa we see: How scholarship from Africa stands to enrich global scholarship. *Africa Journal of Management*, *6*(2), 132–143. https://doi.org/10.1080/23322373.2020.1753493

Basukie, J., Wang, Y., & Li, S. (2020). Big data governance and algorithmic management in sharing economy platforms: A case of ridesharing in emerging markets. *Technological Forecasting and Social Change*, *161*, Article 120310. https://doi.org/10.1016/j.techfore.2020.120310

Bradburn N. M., & Sudman, S. (1979). *Improving interview methods and questionnaire design*. Jossey-Bass.

Chipp, K., Wocke, A., Strandberg, C., & Chiba, M. (2019). Overcoming African institutional voids: Market entry with networks. *European Business Review*, *31*(3), 304–316. https://doi.org/10.1108/EBR-01-2018-0029

Cooke, F. L., Wood, G., & Saunders, S. (2022). Migrants working for migrants: Dependence and discourse in Chinese-owned small commercial businesses in South Africa. *The International Journal of Human Resource Management*, *33*(21), 4308–4333. https://doi.org/10.1080/09585192.2021.1949625

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Devarajan, S. (2011, October 6). Africa's statistical tragedy. *World Bank Blogs*. https://blogs.worldbank.org/en/africacan/africa-s-statistical-tragedy

Fedderke, J. W., de Kadt, R. H. J., & Luiz, J. M. (2001). Indicators of political liberty, property rights and political instability in South Africa: 1935–97. *International Review of Law and Economics*, *21*(1), 103–134. https://doi.org/10.1016/S0144-8188(00)00051-X

Fedderke, J. W., & Luiz, J. M. (2002). Production of educational output: Time-series evidence from socioeconomically heterogeneous populations—the case of South Africa, 1927–1993. *Economic Development and Cultural Change*, *51*(1), 161–187. https://doi.org/10.1086/345575

Fedderke, J. W., & Luiz, J. M. (2007). Fractionalization and long-run economic growth: Webs and direction of association between the economic and the social -- South Africa as a time series case study. *Applied Economics*, *39*(8), 1037–1052. https://doi.org/10.1080/00036840500461899

Fedderke, J., Luiz, J., & de Kadt, R. (2008). Using fractionalization indexes: Deriving methodological principles for growth studies from time series evidence. *Social Indicators Research*, *85*, 257–278. https://doi.org/10.1007/s11205-007-9090-x

Fedderke, J. W., Perkins, P., & Luiz, J. M. (2006). Infrastructural investment in long-run economic growth: South Africa 1875–2001. *World Development*, *34*(6), 1037–1059. https://doi.org/10.1016/j.worlddev.2005.11.004

Gelb, A. H., Ramachandran, V., Shah, M. K., & Turner, G. L. (2007). *What matters to African firms? The relevance of perceptions data* (Policy Research Working Paper No. 4446). The World Bank. https://documents1.worldbank.org/curated/en/980061468010497193/pdf/wps4446.pdf

Jerven, M. (2013). *Poor numbers: How we are misled by African development statistics and what to do about it*. Cornell University Press.

Lehdonvirta, V., Kässi, O., Hjorth, I., Barnard, H., & Graham, M. (2019). The global platform economy: A new offshoring institution enabling emerging-economy microproviders. *Journal of Management*, *45*(2), 567–599. https://doi.org/10.1177/0149206318786781

Liang, C., Hong, Y., & Gu, B. (2024). Monitoring and home bias in global hiring: Evidence from an online labor platform. *Information Systems Research*. https://doi.org/10.1287/isre.2021.0526

Marivate, V. (2024, April 11). *AI as our mirror: Exploring ourselves through AI empowerment | Vukosi Marivate | TEDxPretoria* [Video]. YouTube. https://www.youtube.com/watch?v=TPr6_8frh08

Marquis, C., & Raynard, M. (2015). Institutional strategies in emerging markets. *Academy of Management Annals*, *9*(1), 291–335. https://doi.org/10.5465/19416520.2015.1014661

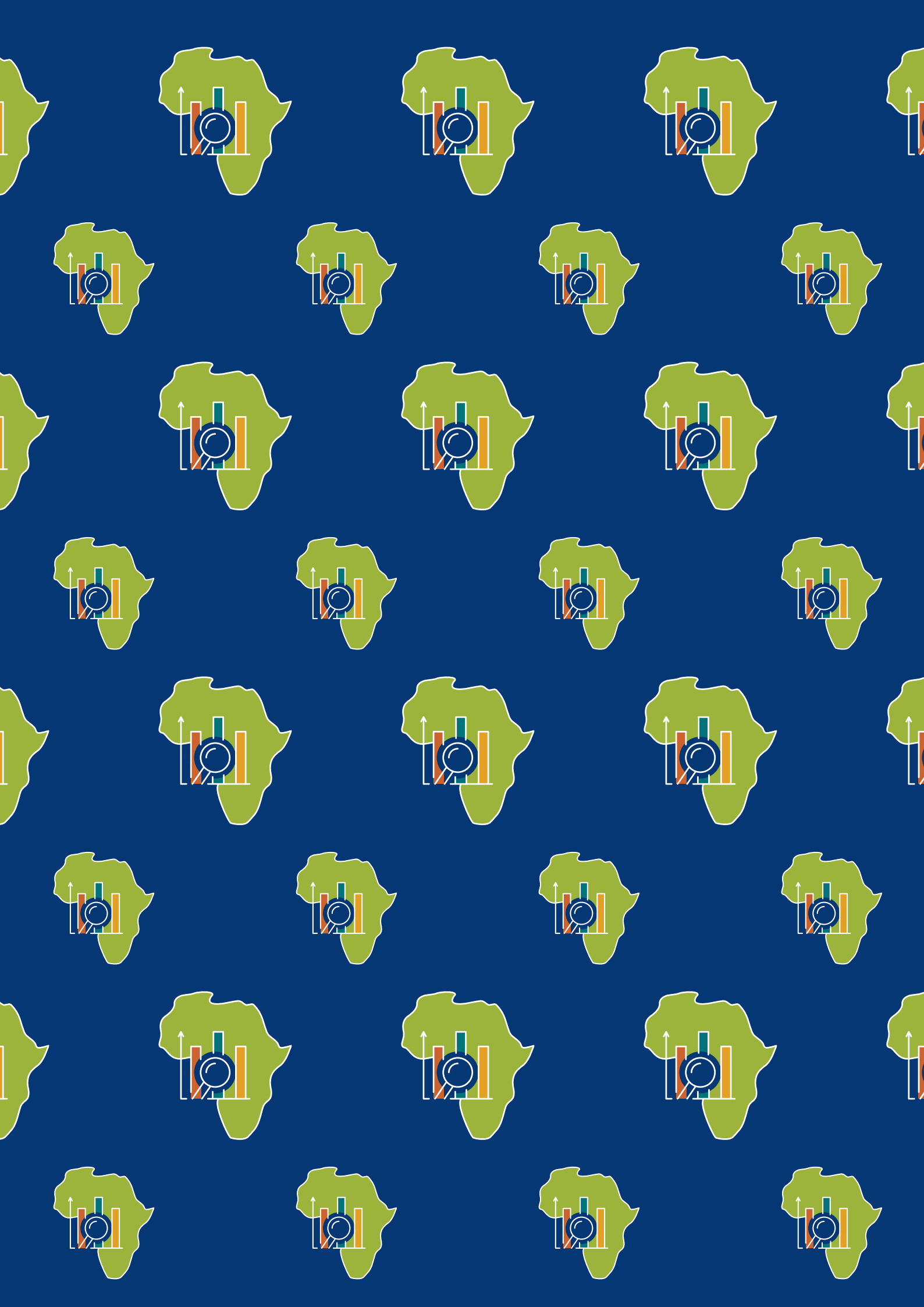
McGuire, J. B., James, B. E., & Papadopoulos, A. (2016). Do your findings depend on your data(base)? A comparative analysis and replication study using the three most widely used databases in international business research. *Journal of International Management*, *22*(2), 186–206. https://doi.org/10.1016/j.intman.2016.03.001

Nair, L. B. (2021). From "whodunit" to "how": Detective stories and auditability in qualitative business ethics research. *Journal of Business Ethics*, *172*, 195–209. https://doi.org/10.1007/s10551-020-04479-4

Ojanperä, S., Graham, M., Straumann, R. K., De Sabbata, S., & Zook, M. (2017). Engagement in the knowledge economy: Regional patterns of content creation with a focus on sub-Saharan Africa. *Information Technologies and International Development*, *13*, 33–51.

Saville, A. (2017). *VISA African integration index: Connecting for growth, economic inclusion & prosperity* (Vol. 4). Gordon Institute of Business Science. https://content.gibs.co.za/cmscontent/media/f4df0ru0/saville_visa-africa-integration-index_vol4_2017.pdf

Wood, G., & Cooke, F. L. (2023). Transient entrepreneurs?: Chinese migrant small commercial businesses in South Africa. *International Journal of Management*, *29*(6), Article 101094. https://doi.org/10.1016/j.intman.2023.101094



Gordon Institute of Business Science
University of Pretoria

26 Melville Road, Illovo, Johannesburg
P O Box 787602, Sandton, South Africa, 2146

011 771 4000 | Acumen@gibs.co.za