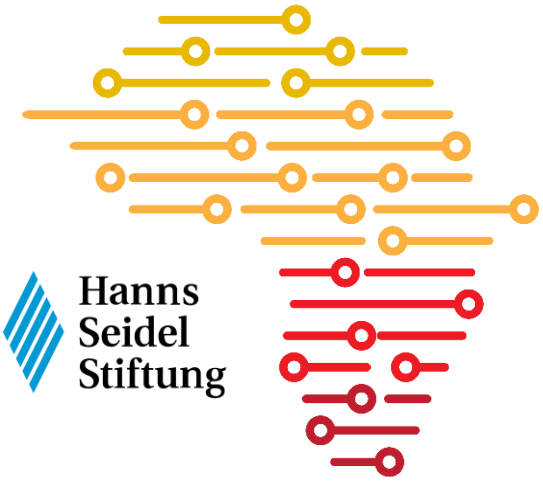
Tayarisha Working Paper Series | Number 2023/016

African Governance in the Digital Age



Illegible data: A counterpoint to the evidence-based approach in policy decision making

Caryn Abrahams

WITS SCHOOL OF GOVERNANCE
UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

TAYA RISHA
African Centre of Excellence in Digital Governance

Illegible data: A counterpoint to the evidence-based approach in policy decision making

Caryn Abrahams

Abstract

The art of governing is closely linked to the statecraft of decision making based on evidence. Population dynamics, social change and the growth of the middle class are some of the data analyses used to shape and influence policy directions and programmatic interventions in particular places. Data analysis is a crucial technology of governance, and it is how place, people and neighbourhoods are known and made legible to the state where the typical goal of the state is the reduction of the 'disorderly, constantly changing social reality beneath it to something more closely resembling the administrative grid of its observations' (Scott, 1998:82). Moreover, this ordering tendency is necessary for decision making that will have the highest impact for the greatest number of people – so there is, indeed, a crucial logic to this approach. Drawing from feminist urban scholars and others, this paper offers a provocation of sorts. It discusses the implications of the homogenising tendency at the heart of data-led approaches to governing. It considers the categories of urban residents and urban life that evade simplification and thus escape meaningful inclusion in policy intervention. This paper does not provoke a discussion of the 'residual/elided/invisible/foreclosed', that is, the illegible, unreadable and unseen in an attempt to diminish the potential and power of data-driven approaches to governance. Rather, it offers a counterpoint, and argues for the other kinds of illegible/unseen/residual data to sit alongside the data considered more legible and credible for decision-making processes. The goal of doing this is so that our knowing, official interests and rational models of policy intervention do not become mechanised or automated, but retain a view on the textual, nuanced and unresolved matters that make up everyday life in everyday contexts.

Introduction

There are many parts of urban life that are illegible – that cannot be read – because there are few measures to read them or to include them in formalised frames of analysis. In their introduction to one of the most influential edited collections about urban life in South Africa – Johannesburg: The Elusive Metropolis (Nuttall and Mbembe, 2008) – eminent scholars outline the intention of the collection which is to celebrate, name, articulate and make visible the kinds of interactions, economic enterprise, identity formation and ways of dwelling that would typically evade categorisation within urban debates in South Africa. They state that many other urban studies

> are more concerned with whether the city is changing along vectors of institutional governance, the deracialisation of service provision, and local politics than about cityness as such [, and…b]ecause they approach the city as a problem to be solved, they are clearly prescriptive. […]They end up mapping an urban social geography of needs, the crucial indexes of which are levels of deprivation. In the process, they underplay many other aspects of city life and city forms. (Mbembe and Nutall, 2008: 3)

This has relevance because there are expressions and definitions of urbanisation that fall outside the scope of analysis – policy or otherwise. There are, as Simone (2008) argues in the same book, different methods of economic engagement, different uses and identification of spaces and bodies in those spaces, different ways of engaging with the state and civic life, and different clandestine activities which 'trip up' the categories that exist to regularise and theorise urbanisation in these cities (Simone, 2008: 87–88).

The distinction between the informal and the formal sectors as it relates to data management is a case in point: the informal sector is seen to need the discipline of regulation and legibility against the unruly nature of its interactions and outcomes, but it is regularised and supported as a result of its legibility as an entity. The gains of the informal sector are systemically illegible within the current taxation system, for example, which precludes enterprises under a certain income threshold to register as VAT (value added tax) vendors (Valodia and Francis, 2020). Thus, it is not the trite notion that the informal sector is somehow deviant in not wanting to have its gains taxed, but rather a systemic issue where large swathes of entrepreneurs in the informal sector are rendered illegible. As Valodia and Francis continue, it is not simply that

the informal sector represents marginalisation such that the intervention relates to service delivery, but more substantively, the informal sector suffers from a lack of systemic inclusion (see also Valodia and Devey, 2012).

There is a self-fulfilling cycle at work here: despite there being plentiful research on the informal economy, since there are no systematised frameworks that make the flows or transactions legible or the interactions regularised particularly within the gross domestic product (GDP), the labour and outputs of the informal, domestic and care sectors tend not to shape decisions about economic policy as a whole (see Floro and Komatsu, 2011; Vaca Trigo and Baron, 2022). Take another example of this same argument related to the GDP calculation: a provocation critiquing how the GDP is constructed (Mushongera, 2017) and how it renders informal, care and domestic work illegible and thus unvaluable to formalised economic measures. This is useful to frame the claim by Vaca Trigo and Barron (2022) that activities of everyday life which do not pass through the market formally, are unvalued and ignored in economic analyses, despite being very central to everyday life because the 'concept of value is reduced to economic value' (Vaca Trigo and Barron, 2022: 7). As Vaca Trigo and Baron (2022: 8) put it '[a]lthough GDP is intended to offer a standardized measurement of output, its use has expanded to become a measure of well-being and the main economic indicator, even though it fails to account for much of what is produced and much of what sustains people's lives'. If the unstructured data were to be rendered usable, they might more centrally shape notions of gross domestic product as including care, unpaid and informal work, and may change taxation systems to allow informal businesses to report on incomes below a threshold (see Valodia and Francis, 2020). This captures what Prahalad termed the 'Fortune at the bottom of the pyramid' (Prahalad, 2006).

These parts of urban life are as equal a part of the fabric of society as those parts of urban life that are measured, quantified and which come to represent value in a city economy. The question is not (just) how can we include these illegible spaces of city living into our frameworks of urban economies but, rather, how our frameworks for analysis, our very tools of understanding, exclude these very vital spaces of urban metabolism. Moreover, as I argue here, the initial stance is that this is not a problem to be remedied necessarily, but rather recognised. This paper, then, does not fit the script of offering recommendations for a social

problem, nor is it based on empirical findings. Instead, it is a provocation which offers some thoughts about the endemic challenge that the use of data in the public sector is beset with. The problem is that the evidence-based planning frameworks that exist cannot read areas of hidden work or enterprise. Nor can they necessarily be designed to.

The everyday exchanges of informal, unpaid care, and domestic work are in some sense 'unseen' because these realities evade the categorised gaze of social ordering of the economies, regularised places of dwelling or the regulated work that we do. Our cities, and the activities that make urban life happen, are known in everyday reality, but these are often out of sight of official indexes of urbanisation trends. Thus the very frames that seek to positively interpret and analyse changes in urban life produce 'elisions', obscuring the varied interactions which are the very textual essence of life in a city. Feminist scholars such as Buckley and Strauss (2016) and Peake and Rieke (2013) for instance, point out the elisions that are produced because of a political economy reading of urbanisation, which centres on a reading of capitalist globalisation, productive capital and economic growth through capital accumulation and infrastructure growth as the basis for urbanisation. They argue that these measures elide the fact that this reading of urbanisation is fundamentally based in heteronormative patriarchal modes of capitalist accumulation and, in so doing, creates a totalising view of urbanisation (Deutsche, 1991) that renders 'feminist, post-modern, queer and other approaches to social difference' invisible (Buckley and Strauss, 2016: 622). There is a further removal of black women's experiences, they note citing hooks (2008), because of systemic exclusion from the formal economy, and not just spaces of capital accumulation. They articulate the task at hand, as a radical feminist reading of the urban, in order to open up its categorisation in a way that

> decentre[s] empirically defined, spatially bounded models in which the materialization of capital flows (the spatial fix) implie[s] a clear divide between public and private space and the analytical privileging of collective reproduction and consumption. […] African American feminist critiques of such white feminist conceptualizations of these divisions between 'home' and 'work' further illustrated the importance of social difference and social location in the production of the urban. (Buckley and Strauss, 2016: 621–622)

Urban feminists add content to descriptions of the 'residual' that falls outside and therefore 'escapes current planetary urbanization frameworks' (Buckley and Strauss, 2016: 619). They urge us to fundamentally rethink what we know about the urban and the frameworks of urbanisation that have come to shape what we know about cities. Similarly, Floro and Komatsu (2011) discuss that the way to remedy turgid analyses of economic productivity related to paid work in the formal economy is to understand the centrality of time-use of women in care, and other unpaid work circumstances, which is often overlooked in urban (un)employment discussions.

Drawing from the vantage point expressed above, this paper offers a provocation which discusses the implications of the homogenising tendency at the heart of data-led approaches to governing. It considers the categories of urban residents and urban life that evade simplification and thus escape meaningful inclusion in policy intervention. This paper does not provoke a discussion of the 'residual/elided/invisible/foreclosed', that is, the illegible, unreadable and unseen in an attempt to diminish the potential and power of data-driven approaches to governances. Rather, it offers a counterpoint and argues for the other kinds of illegible/unseen/residual data to sit alongside the data considered more legible and credible for decision-making processes. The goal of doing this is so that our knowing, official interests, and rational models of policy intervention do not become mechanised or automated, but retain a view on the textual, nuanced and unresolved matters that make up everyday life in everyday contexts.

Conceptual basis of the paper

I take, as the basis of this paper, an argument developed by Scott in the seminal text Seeing Like a State (2008). Scott argues that the art of governing and decision making is integrally connected to the statecraft of developing schemata to make the expanse of social and economic life legible or readable. Data and data analysis is how place, people and neighbourhoods are known and made legible to the state, where the typical goal of the state is the reduction of the 'disorderly, constantly changing social reality beneath it to something more closely resembling the administrative grid of its observations' (Scott, 2008: 82). This ordering tendency is necessary for decision making that will have the highest impact for the

greatest number of people – so there is, indeed, a crucial logic to this approach. Scott argues that there is another reason, apart from rational decision making, that official categorisations, definitions and measures 'stick'. It is because the state relates to society on the basis of these categorisations or schemata. Thus, as he argues, 'categories that may have begun as the artificial inventions of cadastral surveyors, census takers, judges, or police officers can end by becoming categories that organize people's daily experience precisely because they are embedded in state-created institutions that structure that experience' (Scott, 2008: 82–83).

However, similarly to the evocative examples shared at the beginning of this paper, these simplifications are not able to convey or represent the sum of interactions in the slice of life it describes, nor '[a]re they intended to; they represented only that slice of it that interested the official observer' (Scott, 2008: 3). Indeed, if the messiness of society can be alchemised into neat and uniform standardisations, these schemata will become useful, appropriate and amenable to the task that officials have to intervene in society. In other words, if the illegibility is rendered legible, it is more amenable to the effective governing mandate of the state and, conversely, 'an illegible society, then, is a hindrance to any effective intervention by the state, whether the purpose of that intervention is plunder or public welfare' (Scott, 2008: 78).

Data collection regimes, however, are not an alchemy, they are very much a central tenet of the digital age. Nor are the data collection regimes necessarily interested in reducing the amounts of data at the disposal of decision making. Indeed the opposite. Governments collect data, governments author data, and governments own and manage data. This is part of the everyday practices of the state from demographic data through the census to income data through earnings and tax, cross-border travel, residence and use of public services, etc. These form part of government monitoring and evaluation frames, as evidence is fundamental to policy planning and decision making. Commercial data are matched to spatial data and this, in turn, to risk profiles – the more data at the disposal of decision makers, the easier it seems decision making should be. This is precisely the point of data regimes and decision making – the more data we have, theoretically the better decision making would be. But what is better? Who determines it? And is more data better when pragmatic decisions still need to be made? Take, for instance, this example: rational decision making using data considered the Covid-19 temporary relief of distress grant. This intervention was premised on households who were

eligible based on their already receiving any type of grant. The decision was to extend the emergency welfare package beyond only those who were recipients of the child support grant (CSG) because '[w]hilst researchers can analyse survey data on households, no government department collects data at the household level. Social grants are – and have to be – paid to individuals. Relying on CSG supplements alone would encourage the belief that the CSG should be used to support entire households' (Seekings, 2020: 14).

Despite many other probable sources of data, the decision making about the extension of the 'Covid grant' relied only on legible data used as proxy for eligibility for the temporary economic relief support grant, that is, considering those who were also recipients of other grants. Formal workers who worked at companies who already contributed to the unemployment insurance fund (UIF) were eligible to receive emergency temporary relief from the companies who applied for a special portion of the UIF grant facility made available for Covid income support (See Seekings, 2020). Informal workers could not benefit from the companies' UIF grant, but could benefit from the TERS grant only if they were already receiving another kind of grant.

Data regimes are central to urban governance: administrating life in cities is filled with interactions that evade categorisation. Yet, there is a very pragmatic requirement to understanding the trends, the arena of social change and the demands of growing urban populations. Evidenced-based planning is a rigorous and powerful tool that has a particular part to play in decision making. Its value is undisputed. Indeed, it has become central to the language and practice of local and subnational governments. A large part of the results-based management ethos is infused with the very mechanisms of programme and policy evaluation. And with just cause: without data on, for example, the socio-economic status of particular communities, or data on the economic opportunities in particular areas, decision-making and direct interventions about, say, extended public works programmes and how those may benefit communities while implementing small service projects in those areas, would not have reach or impact. Or without income data, the massive social protection grants programmes would not reach the intended beneficiaries. These extend, but are not limited, to knowing how many people are on indigent registers, for example, or how the rate of urbanisation is faring in comparison to the infrastructural challenges that exist.

Evidence matters. Planning based on evidence is quick becoming a non-negotiable part of policy making, across the policy cycle, and in determining the best course of action between various alternatives. As Cairney (2017) notes, 'good policymaking' relates to good theorisation about what will work and why. These summations include evidence-based planning and management. Without this kind of detailed and reasoned analysis (which he rather ambitiously calls policy theories), it is more likely that one would be tempted to 'throw up one's hands and declare that politics and policymaking are dysfunctional, and that policy is made with too much reference to ideology and too little to evidence' (Cairney, 2017: 3).

Yet, evaluating implementation is often treated in a tick-box manner about which there are critiques (see Hill, 2013). For example, seeking mainly to evaluate the performance of implementers or capture the reach of the programme often allows these measures to effectively stand for evaluating how far we have met the overall objective. In Greenhalgh and Russell (2009: 304), the uncritical over-reliance of certain narrow reading of data has resulted in a health policy context in the United Kingdom (UK) where 'a narrowly "evidence-based" framing of policymaking is inherent[ly] unable to explore the complex, context-dependent, and value-laden way in which competing options are negotiated by individuals and interest groups'. Crucially, this is not about the paucity of data – where the solution is perhaps more data collection – but rather the tools of analysis used to interpret the data. Hill (2013) argues that often the kinds of evidence used in the policymaking process, particularly in evaluation, tend not to lend towards accountability, but often get very narrowly framed in terms of indicators of performance. In Khumalo et al. (2022), a rich analysis of African parliamentary processes is provided, looking at the various ways in which evidence is used and institutionalised in decision-making processes. In an edited collection, Mapitsa-Blaser and Mosienyane (2022) argue that, while parliaments are information-dense spaces, the space remains inherently political, concerningly, and for the most part the political executive still shapes, delineates and filters the kinds of information produced and therefore received for analysis. They also show how certain kinds of data are used (and possibly produced) at different time intervals, not primarily to affect the policy cycle, but to enhance political decision making at various levels of granularity during politically pertinent times. A final consideration from their paper for the discussion here is that the kinds of data are profoundly

mediated or 'brokered', and this relates to the politicised nature of the way evidence tends to be used.

To link back to the conceptual framing, it is evident from the discussion about evidence-based planning that the kinds of evidence used, while always useful, can be instrumentalised to further particular ends, most notably politically, within a policy/bureaucratic environment. Not only does this crowd out other competing options or the complex ways in which those options are negotiated in the real world, it also profoundly prescribes the kind of view on the data that is more important, and thus made to be more legible to the political decision maker – this despite data troves at one's disposal. This is, of course, not a new critique; as Desrosières (1998: 3) notes, 'statistical tools allow the discovery or creation of entities that support our descriptions of the world and the way we act on it'. Linking to Scott's central theme in Seeing Like a State, political decision making is part of the homogenising tendency of states to produce the kinds of data or evidence that would allow, justify and otherwise rationalise particular types of intervention. This does more than just allow particular categorisations to simply 'stick'. In creating and using data in this way, governments thus maintain the authoritative power of the state to classify, and maintains itself as the authorising authority (Mbembe, 2001), and also authorises that particular 'gaze' – which is the official and authorised account of reality, or of the problem identification that forms the basis of policy analysis (see Gupta, 2011). As urban planning scholar Fend Flyvbjerg (2003: 319) has so compellingly put it:

> Power determines what counts as knowledge, what kind of interpretation attains authority, as the dominant authority. Power procures the knowledge which supports its purposes, while it ignores or suppresses that knowledge which does not serve it.

Big data, AI, machine learning and policy

Big Data offers profound possibilities for absolutely efficient decision making, near-flawless analysis of trends, projections, forecasting and precisely targeted interventions. The possibilities are well documented in finance, marketing and business sectors more generally (see Dwivedi et al, 2021, as well as for use in the public sector (Rahmanto et al., 2021). Vydra and Klievink (2019) argue that there tends to be techno positivity associated with the

possibilities of AI and big data in the economic sphere. However, while promising to be a game-changer, the possibilities of AI and big data may invite pessimism when it comes to the potential of big data in the policy sphere. They argue this because of a caution about privacy regulation and access to data from a policy perspective. There are, of course, useful lines of inquiry about types of legal and policy frameworks in thinking about its policy relevance, but the focus here is on decision-making through data – in this instance big data and machine learning.

Scholars distinguish the use of big data and data-learning algorithms in the public sphere as those which function as expert systems that support decision making and those that relate to automated decision making (de Carmargo et al., 2018; Dietzman and Duan, 2022; Edwards, et al., 2000). In the former case, the data support decision making in an unstructured manner and, in the latter, data replace it in automated ways (see Di Vaio et al., 2022; Duan et al., 2019). Dietzmann and Duan (2022) also outline the various usages along that spectrum – models for information processing, data acquisition, synthesis and sense-making of data, among other uses. They argue that the use of data depends upon the 'processing capacity of a decision maker' (2022: 5925). There are risks, they add, even in the supportive role that data can play in decision making, since data overload can de facto result in mis-information, where there is uneven definition of problems and over-reliance on external consultants or on support staff. In sum, they argue that the manager's 'decision-making process is influenced by uncertainty, complexity, and equivocality of situations offering potential for AI use' (2022: 5926).

Okwechime et al., (2021) similarly look at public sector decision makers where data guide decisions, which can be applied to predictive models. Here, the scholars make a distinction between the kind of data and whether it is held, authored or used by the state. They consider the use of different kinds of data, including unstructured community-authored data, and argue that open source data that is shared will require similar machine-legible data, but that it offers incredible potential for sources of data (Okwechime, 2021: 41). They discuss the applicability of these large open source datasets to assist decision making, for example, 'in every aspect of urban development such as healthcare, transportation and security'. This, however, does not replace the role of decision makers they argue. Instead, as they put it,

intuition still pays a large role even in big data 'because when addressing complex and unstructured problems, decision makers are influenced by their experiences of other holistic associations as well as time constraints' (Okwechime, 2021: 42).

While this discussion has focused on data use as an expert system to support decision making, there is the potential and probable use of automated decision making to effectively replace human or politically intuitive decision making. Stöger et al. (2021) consider a further distinction here – automated decision making based on a finite and knowable set of inputs used to arrive at an automated once-off decision, where the system is not constantly learning after being 'deployed' or mobilised, and the data and the data processes that are involved in teaching or training AI to constantly (as opposed to once off) recalibrate data inputs, learning to make better decisions over time. Stöger and the co-authors write specifically from a medical-legal perspective, where data sets can be trained to reduce human diagnostic decisions, and to integrate medical decisions with the use of devices that would monitor, predict and correct certain dosages, for instance, or that would consistently monitor medical risk patterns. The article is concerned with what my paper names illegible data, in that it discusses the legal implications of 'problematic' data, which could be incomplete data, unknowns, dummy variables, duplicated or missing data. In any of these cases, the result is unusable data, so the task of analysts and, more specifically, data cleansers, is to remove or account for problematic data, so as to end up with a clean, unambiguous and usable data set. If the aim is certainty, then data that present uncertainty would be considered unusable. However, as I discuss elsewhere, there is value in analysing these categories or the uncertainties presented in data (see Abrahams, 2016).

Cleansing the data from all possible uncertainty is seen as a function of legally modifying software, and creating 'business rules' whereby contradictory data, such as prostrate disease, cannot be related to data inputs where the biological sex is female. These data-cleansing routine processes could result in deletion of those records that present any sort of inaccuracies and contravene the business rules. The article goes on to give compelling examples of how the cleansed data could overlook areas of risk, for example breast cancer in men, or could disadvantage transgendered people.

Where it becomes problematic is the definition of 'business rules' and the automation of biases encoded in the data. The article goes on to argue that seeking data accuracy can rely on particular data that are considered sensitive: 'In some situations, the collection of particularly sensitive data can be necessary for training data in order to prevent discrimination by the system – even if the corresponding data is not used as input' (Abrahams, 2016). Sensitive information could be gendered data, ethnic data or mental health status, which would need to be anonymised. In this case, they argue that European Commission legislation allows for and 'contains an exception to the prohibition on processing sensitive data if processing is strictly necessary for the purposes of ensuring bias monitoring, detection and correction' (Stöger et al., 2021: 6–7). Furthermore, the legislation notes that if anonymised but potentially biased data are used, its fair use needs to be monitored by national authorities as 'the quality of (training) data sets (e.g. representativeness, de-biasing, data and aggregation standards) as main building blocks of AI is vital and that therefore data sets should in the future be auditable by national supervisory authorities' (cited in Stöger et al., 2021: 7).

Clearly, in the medical-legal sphere this has profound consequences for safe medical intervention and decision making, but what about this very principle of data cleansing, the creation of business rules or the use of potentially biased data to train AI machine-learning systems to make decisions that do not have the same 'fair use' or auditable requirement as does medical information in the EU? What if those business rules remove problematic data, which might be a consequence of gender, ethnic or language othering? Or, if an undesignated 'other' category is used because no clear category fits, the business rule could be to cleanse the problematic data in service of a more accurate/legible/unambiguous dataset.

Indeed, this is not a fictitious scenario – the use and weaponising of data tools to remove the existence of undesirable and thus illegible categories of people in the data has been seen to racially and ethnically profile risk categories of travellers, or mortgage or citizenship applicants. Thus large studies of home-loan rejection data in the United States (US) do not look as racially skewed if the 'other' categories are scrubbed. It should be noted that this does not only relate to the 'other' category, but the post-stratification 'other' category, for example, Other – Hispanic; Other Black, mixed-race – where not enough data exist for these

post-stratification categories (see Washington DOH, 2010; Khan et al., 2022). A paper commissioned by the Organization for Economic Cooperation and Development (OECD) indeed focuses on this very issue: how to collect and not purge missing data that may speak to gendered patterns or ethnic or racial data (Balestra and Fleischer, 2018). There is a twin problem, they argue: the collection of sensitive data and the removal of sensitive data. The first has to do with the potential for this data to be weaponised, and the second has to do with the effect of removing all sensitive data, thereby increasing those people's marginalisation and exclusion. Citing others, Balestra and Fleisher (2018) argue that

> it is feared that [sensitive data] may be misused to maintain or deepen power relationships between majority and minority population groups, […a] concern […]particularly heightened for groups who may have experienced ethnic profiling, segregation, genocide and violence […]and especially so in countries where ethnicity-based data were used in the past to provide the basis for discriminatory policies. (Balestra and Fleischer, 2018: 20–21)

The other extreme is similarly pernicious, however. Certain countries, in an attempt to remove bias, engage in a

> "colour- blind" approach that explicitly forbids the collection of any data on collective identities at all. Ethnicity-blindness is supposed to ensure equality of all citizens in front of the State and, consequently, in social life more generally, regardless of ethnic or racial differences. [This] is referred to as "equality through invisibility". (Balestra and Fleischer, 2018: 21).

However, limiting or prohibiting the collection of racial or ethnic data while appearing impartial can 'in practice, disadvantage groups' where there are already structural and economic exclusions or invisibility in the way discussed early in my paper. These 'unknown, unseen, and illegible lives', in the words of Balestra and Fleischer (2018: 21) 'can suffer from the lack of explicit information about their experiences and quality of life, and be disregarded by a democracy that has little knowledge about their demographic presence and the challenges they face'.

What we see here, in my view, is the potential for big data to systemically and systematically render some experiences and existences illegible, while potentially also subjecting these to hyper-surveillance, through creating 'business rules' in data automation that could exclude or maintain hypervigilance on some lives.

The uses of big data for governance and policy decisions have the potential to be deployed (and has been) for surveillance (see Washington DOH, 2010), tracking recidivism (Vydra and Klivink, 2019) or biopolitical, economic ends, and healthcare. However, the more nefarious potential of big data, which is accessible within an ecosystem of data sharing by big business, the state and academia, is not related to privacy issues, but issues of social control. Here there is the potential, using learning-automated systems, to censure particular individuals or groups (see Karpa et al., 2021), to restrict their ability to connect with others, to systematically track political dissenters (see Degli Esposti, 2014) or to allow markets to monitor and control social choice to an even greater extent (see Marks, 2021). As Feldstein (2019: 9) argues that in weak democracies

> these tools offer new possibilities for tracking and intimidating dissenters, monitoring political opponents, and preempting challenges to government power.

It is not far off to imagine the hyper-surveillance of 'risky' communities, gender and biopolitical profiling or targeting ethnic communities for systemic exclusion. The conversation, however, tends to focus on the positive potentialities of dataveillance (Degli Esposti, 2014), rather than the nefarious. As Karpa et al. (2021: 13) note, in the case of China, the potential for data surveillance to exercise social control is 'not yet seen in a negative light by most Chinese citizens', which relates in part to the autocratic nature of the state. However, possibilities do exist for potentially positives outcomes to the ecosystem of data sharing. There are postulated benefits to initiatives like 'Know Your Citizen', an IBM platform for state-owned data that would potentially integrate voter-roll data with health care, and other demographic data (see IMT, 2022), making decision making and intervention targeted and accessible.

Determining structural (ir)relevance

There are three concerns which arise in light of the above discussion. The first and most obvious is the epistemological point – the political gaze effectively comes to determine what is known about a particular social reality because the political gaze authorises and instantiates particular frameworks of knowing, classifying and also categorising reality. The point was raised earlier in the feminist critique of epistemologies of the urban: the issue at hand here is that society becomes legible mainly in political or official ways. Society is thus demarcated on the basis of how we fall into categories of economic activity, political sensibilities, social order or governability. Thus there is the (fairly well-accepted) tendency to smoothen out the rough edges in order to make neat calculations based on interventions. Let's call this tendency approximation (to find a measurable proxy). Approximation is based on certain fairly objective categories, which have to do with determining the need for intervention in an often predetermined way. Take for instance programmes meant to create liveable human settlements: there may be crude approximations on what kind of dwelling might be most appropriate for particular sections of society. It is also related to the 'proxy' used for, in this case, the creation of human settlements. Charlton (2018) notes the homogenising framework that shapes the spectrum of housing interventions, which is rooted in discourses of modernising, regularising and making the messiness more amenable to state intervention. While this is not social control per se, and the ends are arguably righteous – like the need to grade roads in informal settlements so that refuse removal could be easily done, or so that stands could be regularised so that access to land could be formalised – these homogenising tendencies are based on a knowledge frame about urban space that may tend to legitimise particular ways of living and, in so doing, delegitimise others.

The second concern is what lies outside the official or authorised view. Here, I focus on the urban studies scholarship, and also arguments about the nature of the informal and formal sectors. While the fact of intellectual engagement with data and evidence which informs intervention is laudable, what is of concern is the implication of 'restricted and clearly defined economic analysis; […and the] overall categorization on this restricted basis' (Robinson, 2002: 540). (1) In the first instance, there are activities and characteristics of urban life that 'do not show up on the radar' within the categories of urban existence typical of reading or theorising

the Global South (Robinson, 2002: 540). (2) Consequently, and even more damning, because of this rigid framing, categories of urban political economy engagement and that which constitutes urbanisation are rendered 'structurally irrelevant' (Robinson, 2002: 537).

This suggests that the everyday interactions of urban life, the economic and social choices and transactions of those 'not on the radar', or those whose social realties are not reducible to neat categories, or whose everyday existence are illegible to politicised frames of knowing/classification are irrelevant in an otherwise structured system of governance. To put it even more crudely, if these realities are illegible, they are effectively unseen and invisible to a governing gaze and, in that way, the realities that they represent matter not. Or when they are seen, they are hyper-seen, or tracked in ways that ascribe the illegible as a veritable risk category.

The third concern is the effect of illegibility on a social reality or person's/collective's existence: Within a data-led planning framework, a reality that evades categorisation is one that is unseen and thus illegible within governing schema. This illegibility functions often as a de facto and de jure means of de-authorising or rather unauthorising particular realities but, more precisely, it is the tendency for these forms of existence to stand outside the sets of legal authorisation or tools used to regularise society. If there are some realities that fall outside of an authorising epistemological frame, it means that there may be dire consequences for the people and communities they represent if there are no frames to make sense of them. Macklin (2018: 15) argues, in the case of vagrants and homeless people, that the technologies of the state to regularise housing within a very prescriptive and restrictive legal frame means that '[s]ome of these techniques can be so broadly employed that they essentially criminalised status for some urban residents'.

A related argument is the sheer insufficiency of aggregate data to measure micro variables of society and evaluate how these were dealt with, perhaps with some cross-pollination of other methodologies. For instance, Kerr and Wittenberg (2019), in a Working Paper for the World Institute for Development Economics Research, provide a repository of micro-level employment data, which may extend policy and development interventions in South Africa. In relation to informal labour, intimated in the introduction as similarly structurally invisible,

there are indeed attempts to include these forms of labour within structural elements of social protection in South Africa or simply enumeration (see Charman et al., 2017; Kerr et al., 2019). Three such examples are the following: (1) the unemployment insurance fund contribution for minimum wage earners, particularly domestic workers. However, while the possibility of inclusion in the social protection scheme exists, the data from these contributions tend not to form part of income datasets (see Kerr et al., 2019); (2) the ground-breaking Mahlangu Constitutional Court judgement which cited argumentative scholarship of Wits academic Cathi Albertyn who mobilised a feminist jurisprudence lens to shine a spotlight on the rights of informal and particularly female domestic workers to workplace injury compensation, effectively rendering the home in which that worker works a designated work site (see SAFLII, 2020); and (3) a recent civil society lobby policy proposal for mandatory retirement benefits for domestic workers (personal communication, October 2022).

There are several grassroots initiatives that seek to remedy the invisibility in official statistics and data. These initiatives are particularly designed to create data outputs that are usable to policy makers because they are formatted in ways that can merge with existing structured datasets, and the authorship is in close collaboration with the state (see ASTON, 2020). The data initiatives range from providing unstructured data through narratives, datasets and maps (see youthmappers.org). Indeed, there is an entire counter-mapping movement that is committed to reversing the spatial determinants of physical infrastructure, and so creating a different contextual narrative to the official ones. These are seen not as collaboration, but rather resistance (Ortega et al., 2018).

What does this discussion mean for data-led, evidence-based policy making?

1. If it falls outside the gaze, it is left to 'politics' [but the gaze is politics], so that our interventions in society do not always (and only) follow what can be measured, ordered and controlled. This relates to what Dietzman and Duan (2022: 5923) call a 'quantification bias' where only those things that can be measured fall within the system of analysis.
2. There is a need to resist crude categorisation for assessment and evaluation frameworks.

3. Critical scholars need to temper the intellectual repository and make space for uncertainty. 'Researchers, policy makers and urban activists can practice ways of seeing and engaging urban spaces that are characterized simultaneously by regularity and provisionality' (Simone, 2010: 69).
4. Radical epistemological activism requires us to push the boundaries/possibilities for other kinds of illegible/unseen/residual data to sit alongside the data considered more legible and credible for decision-making processes.
5. And finally, we must commit as critical scholars to correct the tendency to exclude, marginalise and instantiate feelings of anomie and irrelevance [insignificance] within, those whose realities are too unwieldly to place within a neat set of categories.

For the most part, the earlier discussion centred on data use in general – and not on the quantum of data. The subsequent section dealt with the applicability of big data, automatic data analyses and AI learning systems to the policy environment and its connection to the conceptual framing of structural illegibility, developed above.

Conclusion

This paper considers the elements of society and urban life that remain hidden in policy decision making within a data-led evidence planning framework, which ranged from unstructured to structured, monitoring evidence and big data. The first provocation dealt with the informal economy and other informal forms of urban existence. The tendency for illegible stories and urban existences to be rendered illegible and delegitimate in official terms, where the urban governing frame is particularistic and epistemologically rigid, has a tautological effect – only 'seeing' within that particular frame. This consequently tends to lead the state to be prescriptive in a way that is self-referential. This elides the possibilities, we can infer, to make full use of the potential within these unseen interactions to remake civil culture, to build just and empathetic communities or to nurture viable enterprise.

Although these elided realities are treated as if they do not matter, as though 'they bring little to the table of prospective collaboration and participate in few of the mediating structures that deter or determine how individuals interact with others' their 'seemingly minimalist

offering – bare life – is somehow redeemed. It has allowed innumerable possibilities of combination and interchange that preclude any definitive judgement of efficacy or impossibility' (Simone, 2010: 89).

There is a cautionary tale here: the second provocation of the paper. The use of data can be politicised and, even when political decision making is removed, biases can be built into the system of data analytics. These can be used to systematically exclude particular 'messy' unreadable data, creating what can be called secondary marginalisation of some individuals or groups. Not only does this potentially affect vulnerable people, such as the homeless, gender atypical identities or informal women workers, in that they are illegible to potential positive policy intervention, but it also raises the precarity of these existences which can alternatively be hyper-surveilled, which is the third provocation of this paper. The concluding provocation is what the three implications of structural irrelevance and unseen/cleaned/invisible data would mean potentially for whole communities of practice, precarious work, gendered labour and oppressed or marginalised categories of people – (1) homogenisation; (2) rendering invisible; and (3) rendering illegal.

On the one hand, illegible elements are rendered invisible or unreadable because of the evidence-based planning approach that relies on data that can be retrieved and where that data can be analysed in a homogenised fashion. On the other hand, there could be a concerned effort to render these elements hyper-visible to effect social control. And overlaid with the categorisation that seeks to render human beings, urban activity or data structurally irrelevant, there may be nefarious probabilities, particularly if machine learning continuously builds language models and decisional frames within this context. There are yet unanswered questions about the role of power in all this as incidentally or intentionally framing political decisions based on big data analysis. Of course, there is potential to mitigate the probable negative consequences of this unholy trinity of big business, the state and academia which creates and shares data that exist in a closed loop – at best perpetuating policy outcomes, at worst, the status quo or fascist rule. It is important to retain the principle of data-driven approaches supporting and not replacing decision making. The goal of doing this is so that our knowing, official interests and rational models of policy intervention do not become mechanised or automated, but retain a view on the textual, nuanced and unresolved matters

that make up everyday life in everyday contexts. Civil society intervention, the protection of data sources and the possibilities to include these in decision making need to be protected so that no person is rendered structurally irrelevant.

References

Abrahams C. 2016. The national project: Uncertainty and glimmers of hope. In Lefko-Everett K., Govender R. and Foster D. (eds). *Rethinking reconciliation: evidence from South Africa*. Cape Town: HSRC Press. http://hdl.handle.net/20.500.11910/10949 South African Reconciliation Barometer, Institute of Justice and Reconciliation, HSRC Press, Cape Town.

ASTON. 2020. Baseline Study: Collaborative Research to Inform Network Activities (Baseline Study No. 1: 168). ASTON Digital Transition for Sustainable and Inclusive Cities. https://aston-network.org/wp-content/uploads/2021/05/Aston-BaselineStudy-VA-210430-JB.pdf

Balestra C. and Fleischer L. 2018. Diversity statistics in the OECD. How do OECD countries collect data on ethnic, racial and indigenous identity? *Working Paper SDD/DOC(2018)9*. Stats and Data Directorate Working Paper Series: 1–55. OECD. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=SDD/DOC(2018)9&docLanguage=En

Buckley M. and Strauss K. 2016. With, against and beyond Lefebvre: Planetary urbanization and epistemic plurality. *Environment and Planning D: Society and Space* 34(4): 617–636. https://doi.org/10.1177/0263775816628872

Cairney P. 2017. *The Politics of Evidence-Based Policy Making* (No. 1; pp. 1–9). Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.268

Charlton S. 2018. Spanning the spectrum: Infrastructural experiences in South Africa's state housing programme. *International Development Planning Review* 40(2): 97–120. https://doi.org/10.3828/idpr.2018.6

Charman A.J.E., Petersen L.M., Piper L.E., Liedeman R. and Legg T. 2017. Small Area Census Approach to Measure the Township Informal Economy in South Africa. *Journal of Mixed Methods Research* 11(1): 36–58. https://doi.org/10.1177/1558689815572024

de Camargo Fiorini P., Roman Pais Seles B.M., Chiappetta Jabbour C.J., Barberio Mariano E. and de Sousa Jabbour A.B.L. 2018. Management theory and big data literature: From a review to a research agenda. *International Journal of Information Management* 43: 112–129. https://doi.org/10.1016/j.ijinfomgt.2018.07.005

Degli Esposti S. 2014. When big data meets dataveillance: The hidden side of analytics. *Surveillance and Society* 12(2): 209–225.

Desrosières A. 1998. *The politics of large numbers: a history of statistical reasoning*. Harvard University Press.

Di Vaio A., Hassan R. and Alavoine C. 2022. Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting and Social Change* 174: 121201. https://doi.org/10.1016/j.techfore.2021.121201

Dietzmann C. and Duan Y. 2022. Artificial Intelligence for Managerial Information Processing and Decision-Making in the Era of Information Overload. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2022.720

Duan Y., Edwards J.S. and Dwivedi Y.K. 2019. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management* 48: 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Dwivedi Y.K., Hughes L., Ismagilova E., Aarts G., Coombs C., Crick T., Duan Y., Dwivedi R., Edwards J., Eirug A., Galanos V., Ilavarasan P.V., Janssen M., Jones P., Kar A.K., Kizgin H., Kronemann B., Lal B., Lucini B., … Williams M.D. 2021. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 57: 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Edwards J.S., Duan Y. and Robins P.C. 2000. An analysis of expert systems for business decision making at different levels and in different roles. *European Journal of Information Systems* 9(1): 36–46. https://doi.org/10.1057/palgrave.ejis.3000344

European Commission. Directorate General for Justice and Consumers. 2021. *Guidance note on the collection and use of equality data based on racial or ethnic origin*. Publications Office. https://data.europa.eu/doi/10.2838/06180

Feldstein S. 2019. The Global Struggle over AI Surveillance. Carnegie Endowment for International Peace. https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847

Floro M.S. and Komatsu H. 2011. Gender and Work in South Africa: What Can Time-Use Data Reveal? *Feminist Economics* 17(4): 33–66. https://doi.org/10.1080/13545701.2011.614954

Flyvbjerg B. 2003. Rationality and Power. In Campbell S. and Fainstein S.S. (eds). *Readings in Planning Theory*. 3rd ed: 318–329. Blackwell.

Greenhalgh T. and Russell J. 2009. EVIDENCE-BASED POLICY MAKING a critique. *Perspectives in Biology and Medicine* 52(2): 304–318.

Gupta D.K. 2011. *Analyzing public policy: concepts, tools, and techniques*. 2nd ed. CQ Press.

Hill M.J. 2013. Conclusion: Evaluation and Accountability. In *The Public Policy Process*. Sixth edition: 295–318. Pearson.

hooks b. 2000. *Feminist Theory: From Margin to Centre*. Cambridge: South End Press.

IMT. 2022. Know Your Citizen. A Critical Foundation for EGovernment. https://www.imt.ca/know-your-citizen/

Karpa D., Klarl T. and Rochlitz M. 2021. Artificial Intelligence, Surveillance, and Big Data. Working Paper Forthcoming in Diginomics Research Perspectives: 1–29.

Kerr A., Wittenberg M. and UNU-WIDER. 2019. *Earnings and employment microdata in South Africa* 47(2019). UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2019/681-4

Khan A.T., Gogarten S. M., McHugh C.P., Stilp A.M., Sofer T., Bowers M.L., Wong Q., Cupples L.A., Hidalgo B., Johnson A.D., McDonald M.-L.N., McGarvey S.T., Taylor M.R.G., Fullerton S.M., Conomos M.P. and Nelson S.C. 2022. Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. *Cell Genomics* 2(8): 100155. https://doi.org/10.1016/j.xgen.2022.100155

Khumalo L.S., Mapitsa C.B., Morkel C., Masvaure S. and Semela M.K. (eds.). 2022. African Parliaments Volume 2: Systems of Evidence in Practice. 1st ed. African Sun Media. https://doi.org/10.52779/9781991201539

Macklin A. 2018. The Return of Banishment: Do the New Denationalisation Policies Weaken Citizenship? In Bauböck R. (ed.). *Debating Transformations of National Citizenship*: 163–172. Springer International Publishing. https://doi.org/10.1007/978-3-319-92719-0_31

Mapitsa-Blaser C. and Mosienyane T. 2022. A Synthesis of Parliamentary Institutions and Evidence Use. In *African Parliaments: Systems of Evidence in Practice* 2: 321–337.

Marks M. 2021. Biosupremacy: Big Data, Antitrust, and Monopolistic Power Over Human Behavior. University of New Hampshire Franklin Pierce School of Law; Senior Fellow and Project Lead of POPLAR, Petrie-Flom Center, Harvard Law School. 55 U.C. *Davis Law Review* 513. Harvard. https://ssrn.com/abstract=3695373

Mbembe A. 2001. Of Commandment. In *On the Postcolony*: 24–45. University of California Press.

Mbembe A. and Nuttall S. 2008. Introduction: Afropolis. In *Johannesburg: The Elusive Metropolis*: 1–36. Duke University Press. https://doi.org/10.2307/j.ctv125jhd1

Mushongera D. 2017. Beyond GDP in assessing development in South Africa: The Gauteng City-Region Socio-Economic Barometer. *Development Southern Africa* 34(3): 330–346. https://doi.org/10.1080/0376835X.2017.1308246

Nuttall S. and Mbembe A. (eds.). 2008. *Johannesburg: The Elusive Metropolis*. Duke University Press. https://doi.org/10.2307/j.ctv125jhd1

Okwechime E., Duncan P.B., Edgar D.A., Magnaghi E. and Veglianti E. 2021. Big Data: An Introduction to Data-Driven Decision Making. In Magnaghi E., Flambard V., Mancini D., Jacques J. and Gouvy N. (eds.). *Organizing Smart Buildings and Cities* 36: 35–46. Springer International Publishing. https://doi.org/10.1007/978-3-030-60607-7_3

Ortega A.A.C., Martinez Ma. S. M., Dayrit C. and Saguin K.K.C. 2018. Counter–Mapping for Resistance and Solidarity in the Philippines Between Art, Pedagogy and Community. In T*his Is Not an Atlas*: 144–151. transcript Verlag. https://doi.org/10.1515/9783839445198-020

Peake L. and Rieker M. (eds). 2013. *Rethinking feminist interventions into the urban*. Abingdon, Oxford: Routledge: 1–23.

Prahalad C.K. 2006. *The fortune at the bottom of the pyramid.* Dorling Kindersley.

Rahmanto F., Pribadi U. and Priyanto A. 2021. Big Data: What are the Implications for Public Sector Policy in Society 5.0 Era? IOP Conference Series. *Earth and Environmental Science* 717(1): 012009. https://doi.org/10.1088/1755-1315/717/1/012009

Robinson J. 2002. Global and world cities: A view from off the map. *International Journal of Urban and Regional Research* 26(3): 531–554. https://doi.org/10.1111/1468-2427.00397

SAFLII. 2020. Mahlangu and Another v Minister of Labour and Others (CCT306/19) [2020] ZACC 24; 2021 (1) BCLR 1 (CC); [2021] 2 BLLR 123 (CC); (2021) 42 ILJ 269 (CC); 2021 (2) SA 54 (CC) (19 November 2020), (2020) (testimony of SAFLII). http://www.saflii.org/za/cases/ZACC/2020/24.pdf

Scott J.C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press. https://doi.org/10.2307/j.ctvxkn7ds

Seekings J. 2020. Bold promises, constrained capacity, stumbling delivery: the expansion of social protection in response to the Covid-19 lockdown in South Africa. *CSSR Working Paper No. 456: 1–49*. Centre for Social Science Research. http://www.cssr.uct.ac.za/sites/default/files/image_tool/images/256/Publications/WP456Seekings.pdf

Simone A. 2010. People as Infrastructure: Intersecting Fragments in Johannesburg. In *Johannesburg: The Elusive Metropolis*: 68–90. Duke University Press. https://doi.org/10.2307/j.ctv125jhd1

Stöger K., Schneeberger D., Kieseberg P. and Holzinger A. 2021. Legal aspects of data cleansing in medical AI. *Computer Law & Security Review* 42: 105587. https://doi.org/10.1016/j.clsr.2021.105587

Vaca Trigo I. and Baron C. 2022. Decentring GDP: well-being, care and time. Project Documents: *Social Inequality and the Future of Workers in Latin America in the Context of the Post-Pandemic Recovery* (LC/TS.2022/80). Economic Commission for Latin America and the Caribbean (ECLAC). ECLAC.

Valodia I. and Devey R. 2012. The Informal Economy in South Africa: Debates, Issues and Policies. *The Journal of Applied Economic Research* 6(2): 133–157. https://doi.org/10.1177/097380101200600203

Valodia I. and Francis D. 2020. Tax and the informal economy: Lessons from South Africa. In in Chen M. and Carré F. (eds). *The Informal Economy Revisited: Looking Back, Thinking Forward*: 92–97. Routledge.

Vydra S. and Klievink B. 2019. Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly* 36(4): 101383. https://doi.org/10.1016/j.giq.2019.05.010

Washington Department of Health (DOH). 2010. *Guidelines for Using Racial and Ethnic Groupings in Data Analyses.*

https://doh.wa.gov/sites/default/files/legacy/Documents/1500//RaceEthnGuidelines.pdf