The Bottom of the Data Pyramid: Big Data and the Global South

PAYAL ARORA[1]
Erasmus University Rotterdam, The Netherlands

To date, little attention has been given to the impact of big data in the Global South, about 60% of whose residents are below the poverty line. Big data manifests in novel and unprecedented ways in these neglected contexts. For instance, India has created biometric national identities for her 1.2 billion people, linking them to welfare schemes, and social entrepreneurial initiatives like the Ushahidi project that leveraged crowdsourcing to provide real-time crisis maps for humanitarian relief. While these projects are indeed inspirational, this article argues that in the context of the Global South there is a bias in the framing of big data as an instrument of empowerment. Here, the poor, or the "bottom of the pyramid" populace are the new consumer base, agents of social change instead of passive beneficiaries. This neoliberal outlook of big data facilitating inclusive capitalism for the common good sidelines critical perspectives urgently needed if we are to channel big data as a positive social force in emerging economies. This article proposes to assess these new technological developments through the lens of databased democracies, databased identities, and databased geographies to make evident normative assumptions and perspectives in this under-examined context.

Keywords: big data, Global South, bottom of the pyramid, biometric identities, inclusive capitalism, crowdsourcing, database, democracy

Introduction

"Big data" is a misnomer. While the field is relatively young, much thought has already been put into critiquing the term, particularly equating size with representation. Today, it is hard to argue against the understanding that a dataset may be impressively large, but not necessarily random or reflective of a global and diverse public. Context continues to matter, although it is much more challenging to apply when big data is used in varied and unpredictable ways. Power relations continue to be structured within

these databased ecologies, framed as the 'big data divide" (Andrejevic, 2014) or what boyd and Crawford (2012) have noted, the divide between "the big data rich" (organizations that can generate, purchase, and store large datasets) and the "big data poor" (those excluded from access to the data, expertise, and processing power; p. 674). While focusing on the "divide" in the big data realm is noteworthy, what is problematic is the usurpation of the digital divide discourse to encapsulate power struggles among a generic marginalized populace, as if we live in a global data commons.

The fact remains that while the majority of the world's population reside outside the West, we continue to frame debates on surveillance, privacy, and net neutrality and the demand for alternative models and practices to sustain the digital commons by mostly Western concerns, contexts, and user behaviors. As Udupa (2015) argues, "it is important to widen the lens of media research beyond the western worlds," by conceptualizing new media developments via the lens of "context, variation and power" (p. 2). Perhaps a decade ago it was legitimate to argue that much of this marginalized demographic was not connected to the digital realm and, thereby, could not be incorporated into the contemporary debate, relegating them to development studies experts. Since then however, with the exponential growth of mobile technologies in even the most disadvantaged contexts, along with liberalization policies, and public–private sector commitments to provide connectivity to even the most deprived areas of the Global South, this is no longer a valid argument. It is not only the usual suspects such as China and India taking over the digital sphere, but even countries such as Saudi Arabia and Myanmar. For instance, in Myanmar the shift has been from a mere 1% of its population being online a few years ago to almost 50% by the end of this year (GSMA, 2014). Or, take the case of the sub-Saharan African region where the number of mobile phone subscribers has increased at a rate of 18% annually since 2007, reaching 253 million in 2013. In fact, it is forecasted that by 2020 the majority of geolocated digital data will come from these emerging economies.

The majority of this population, however, continue to live below US$2 a day and come with diverse cultural modes of being, much of which remain a black hole to Internet-savvy scholars and the public at large. C. K. Prahalad, a neoliberal guru in business studies, coined the term "bottom of the pyramid" (BoP) to describe this roughly four billion people (2013). He argued that it was time to reframe this populace as "consumers" instead of "beneficiaries," moving away from persistent colonial perspectives driven by white guilt and paternalism. This would be a win-win solution for both the market and the state, where common good sits side by side profit making. This viewpoint gained a further boost with the rise of Web 2.0 technologies and the cultural shift in perceiving users as cocreators and masses with collective intelligence and wisdom (Surowiecki, 2004). Hence, we are called on to envision the poor as future digital data consumers and agents of change. This fits well with the current call by Couldry and Powell (2014) to situate the "notion of voice" at the center of analyzing the cultures of datafication. The challenge in reconceptualizing and reconfiguring new media productions driven by algorithmic power is in assessing how the authoritarian data regimes structured from above intersect with reflexive and resistive practices from below.

This article argues that while there is much skepticism and caution on the social impact of big data in the West, there is a bias in framing big data as an instrument of empowerment in the Global South. Discourses around big data projects in the Global South have an overwhelmingly positive

connotation. For instance, India is largely celebrated for its biometric identity project that links personal identities with welfare schemes (Parker, 2011). Ushahidi, a social entrepreneurial initiative, is seen as a groundbreaking crowdsourcing platform, providing real-time crisis maps for humanitarian relief efforts in Africa, Haiti, and elsewhere (Rai, 2010). And it is hard to critique Facebook as it steps up to provide free access to select sites such as Wikipedia in response to poor children's struggle to gain access to knowledge in places like Nigeria (Eagle, 2015). However, if we are to channel big data as a positive social force in these contexts, we need to ask critical questions of these projects, decoupling moral and political economies. Hence, this article proposes to analyze these new media developments through the lens of *databased democracies, databased identities*, and *databased geographies* to make evident underlining normative assumptions and perspectives in these underexamined contexts. The intent here is to open up debate in the areas of big data policy and practices in the Global South and give a more global dimension to big data studies.

Databased Democracy

With big data projects emerging in the Global South in the name of the poor, it is worth asking whether we are experiencing a *databased democracy* today. Lyon (2013) calls this "the caring panopticon," incorporating aims of care rather than merely discipline, allowing us to take on a broader perspective on surveillance for the common good than the typical Foucaultian lens. Here, the state positions itself as a servant of the public, and shared data is its democratic enabler. In contrast, we have the governmentality of big data, establishing its power and control through these new technological regimes of practice. While equity is at the center of these data initiatives to signal transparency and trust among citizens, it may continue to remain a form of self-marketing for governments unless and until these initiatives are aligned with genuine restructuring by the state for an open political process (Taylor & Schroeder, 2014).

Currently, open data initiatives are the vogue in the Global South. Chile, Bahrain, Kenya, Brazil, and several other countries have openly shared hundreds of datasets on demographics, public expenditures, and natural resources for public access to foster smart and livable cities for their citizens (Linders, 2013). However, information only becomes knowledge and power for the social good when data is put to uses conducive to its diverse and underrepresented citizens (Hilbert, 2013). Rather than stay suspended in the dichotomous nature of this dialogue on databased democracies, this section explores the complexities around two big data trends that have gained tremendous attention as instruments of the common good. They are (1) biometric identities (e.g., India) and (2) crowdsourcing apps for development (e.g., Africa).

Biometric Identities, Anonymity, and Colonial Lineage

The ambitious Biometric Identity Project in India promises a unique identification number (UID) to each of its citizens through a consolidation of 12 billion of its citizens' fingerprints, 2.4 billion iris scans, and 1.2 billion photographs (Sarkar, 2014). While the West already has existing biometric identity systems, it is nowhere near India's scale and scope. This project is the brainchild of the technology

entrepreneur Nandan Nilekani, who cofounded and built the multibillion-dollar outsourcing company Infosys prior his appointment by the government to head this project.

This project has received much media attention, with discourse leaning toward the empowerment of the marginalized. R. S. Sharma, secretary of the Department of Electronics & Information Technology in India, recently declared that "digital India is not for rich people . . . it is for poor people" (Toness, 2014, para. 3). Credible sources such as the BBC provide further endorsement where, "[the poor] with no proof to offer of their existence will leapfrog into a national online system, another global first, where their identities can be validated anytime anywhere in a few seconds" (Rai, 2013, para. 2). While the West appears to be moving away from the convergence of datasets due to privacy laws, constitutional rights, and public concern, these very initiatives in the Global South are celebrated.

This contradictory attitude stems from the fact that the majority of India's citizens lack passports or other forms of identity, making it difficult to disseminate welfare benefits to the masses. Welfare benefits valued at approximately $60 billion are siphoned off by middlemen using fake identities, leaving the anonymous poor helpless in the face of such acts. Hence, the UID project is positioned as a crusade against corruption. Being against such big data initiatives becomes synonymous with being against the poor. While this project can be a democratic enabler and a force for social good, we need to ensure that we are asking the right questions of these institutional and governing technologies.

Already, a few critical voices have come out against the project, the most prominent among them is the Nobel Prize winner Amartya Sen. He argues that these surveillance structures come with high social costs for vulnerable groups, as the programs entail a tremendous loss of privacy and possible criminalization of those not conforming to the state (Sarkar, 2014). Without strong constitutional personal data protection laws, this system can lead to a number of human rights violations. It is even more disturbing to find that this project has its roots in a 1992 government campaign to deport undocumented Bangladeshi immigrants through the tracking ability of the biometric identity database. Pötzsch (2015) argues that these digital techniques of enlisting the body, or what he terms as "ibordering," individualize the border by attaching itself to mobile bodies through the technical and the biological. Here, the body is the border.

Secondly, while it promises to be a voluntary act, this new digital identity is tied to welfare benefits, participation in public religious pilgrimages, and other societal events. Clearly, people are operating within structured power relations that they are often powerless to contest. Thirdly, there is no such thing as infallibility in authentication. When compromised, they are even harder to re-secure compared to digital signatures (Sarkar, 2014). For instance, using a high-resolution camera, an iris can be captured remotely without a person's knowledge. As for data breaches, there are few contingency plans in spite of the fact that this is a real possibility. Recent examples reek of state vulnerability such as the U.S. government's experience of a breach of sensitive military data and the hacking of Israel's citizen database. German hackers have proved that they can deceive fingerprint scanners and even iris scanners. Furthermore, subjects can and do subvert such programs by mutilating their bodies (Breckenridge, 2005).

Fourthly, pertaining to such a demographic, we need to account for what Magnet calls "high tech racism" (2011, p. 28). Certain bodies are more "unreadable" than others. For example, farmers and construction workers often have worn out fingerprints that scanners reject due to their "low quality attributes." Similarly, a common malady among the poor are cataracts that leave irises not conducive to digitization. This provides exclusion of the very subjects that these programs are intended to include. As Van der Ploeg argues, the UID becomes "a machine-readable witness against the subject" (2005, p. 113). It furthers the exploitation of these people through protracted bureaucratic procedures due to new layers of denial, expanding the cycle of poverty and exclusion.

Rao (2013) conducted an urban case study on the Delhi homeless to empirically address whether the UID project furthered the neglect or empowered this chronically invisible populace. By focusing on implementation, Rao was able to capture how these databased tools perpetuate conventional citizen–state relations and advance marginalization through the exercise of "class arrogance, social indifference, and corruption" (2013, p. 72). While the datafication of "established" citizens was successfully executed, the homeless faced numerous obstacles, including the social worker's ability to mediate on their behalf with the state, their access to patronage, the agreeability of their "body" to digitization due to the consequences of living on the streets, and the cooperation of the IT specialists allowing for retrials in scanning fingerprints and irises. What this reiterated is that this system bears "the seed for future discrimination against all those who do not fit into the electronic mould of 'middle-class machines'" (p. 75).

Lastly, far from the claim that these initiatives are novel and unprecedented, we must recognize that these surveillance systems have deep roots in colonial identification practices. During the mid-19th century, to police their colonies, the British instituted biometric surveillance through fingerprinting (Owen, 2014). The fear of an uprising was a constant motivator to identify and track their "unruly" subjects. Here, "unquestioned authority went hand in hand with pervasive fears of being deceived by the populace" (Sengoopta, 2003, p. 204). Interestingly, even during the colonial days, fingerprints served as proof of identity to access services such as the pension system.

Hence, before we are quick to celebrate these initiatives, we must recognize that these databased techniques of democracy are assemblages of institutions, policies, histories, cultural practices, and situational contexts that play out in a complex unison to materialize and articulate plural realities of governance. In the datafication of the body, there is an innate assumption of harmony among the state, the data, and the body. This mythification of harmony serves as a serious barrier to thoughtful, reflexive policy and practice. Furthermore, while several postcolonial nations have gained newfound confidence and a strong national identity in current times, much of their contemporary institutions are products of colonialism on which projects such as these gain a foothold. By remaining unaware of such structures, we perpetuate colonial regimes of surveillance. Instead, this should serve as an opportunity to rethink and restructure traditional systems, if we are to have a vibrant postcolonial democracy.

Bottom of the Pyramid, Crowdsourcing, and Inclusive Capitalism

When analyzing the Global South, one of the most influential paradigms framing their poor is the bottom of the pyramid (BoP) rubric. By embracing the BoP perspective of the poor as empowered consumers, we are in fact marketizing the poor. It seems today that the pathway to democracy is through inclusive capitalism, the extending of the market system to the poor. BoP economies are on the rise across the Global South. By 2011, a total of 439 BoP initiatives were recorded in nine sub-Saharan African countries, targeting the needs of people living on incomes of less than $2 a day in multiple domains, including health and education to information technologies and energy (Blowfield & Dolan, 2014). Informal economies of the poor are brought into the fold by this neoliberal effort. Several corporations see the virtue of this perspective and are vigorously experimenting on "doing good," simultaneously gaining the first-mover advantage among this future consumer base. This entails the conflating of moral and market objectives, where the previously "unusable" poor become a viable and virtue-laden market.

In parts of Africa a number of corporations, not coincidentally some with colonial legacies such as Unilever and Cadbury, transform participation by the poor into a new form of market research. As a Unilever representative remarked as the company promoted good hygiene practices in these regions, "Of course, when we are talking about toothpaste, it happens to be Close-Up . . . when we are talking about soap, it happens to be Lifebuoy" (Dolan & Roll, 2013, p. 14). Marketing literature has long proved that once you shift consumers' behavior in a particular realm, you are well placed to gain their loyalty across an entire category of products. This is no different with the IT industry. Currently, Facebook through its Free Basics platform has promised to provide free access to select websites to the poor in emerging economies. In doing so, Facebook becomes the Internet to this substantial BoP user base. These are by no means a random set of websites, but rather sites that are globally popular such as Wikipedia so that, as Facebook states, poor children can access and learn at no cost and gain an education. Net neutrality takes a backseat in the name of doing good and gives Facebook a unique vantage into the databased behavior of the BoP populace. Wikipedia serves as a Trojan horse, paving the path for Facebook's monopoly among the Global South's data-driven subjects. Democracy of information and select IT brands are indelibly tied together through such efforts.

Big data has opened up new forms of social entrepreneurship in emerging markets, promising social equity hand-in-hand with corporate profit. This technocratic strategy confronting poverty revives the "new frontier" rhetoric of the Internet, where anyone and everyone can benefit from these opportunities— a social, profitable, and moral digital commons (Arora, 2014). Here, "poverty capital" (Roy, 2010) goes hand-in-hand with "big data capital." As Dolan and Roll (2013) argue,

> these initiatives create BoP economies through a set of market technologies, practices, and discourses that render the spaces and actors at the bottom of the pyramid knowable, calculable, and predictable to global business and . . . technologies extend new forms of market governance over the informal poor, reconfiguring their habits, social practices, and economic strategies under the banner of poverty reduction. (p. 124)

The critique of these BoP models, however, confront a number of big data initiatives that have proved to be empowering such as Ushahidi, a crowdsourcing platform (Philip, Irani, & Dourish, 2012). This site draws data from different ICT channels into real-time crisis maps to assist humanitarian relief efforts. In 2010 Ushahidi launched a crisis map in a mere four days after the Haitian earthquake, well before other agencies could respond. In fact, crowdsourcing as a democratic instrument has gained prominence in the Global South, employing collective power to address social challenges, and has become a buzzword in the humanitarian sector (Hellström, 2015).

Several applications of big data for development have emerged since then: an application that tracks location-based data (e.g., GPS data from mobile devices) and nature-related data (e.g., weather) to help farmers; Nextdrop, a crowdsourcing app that, for a nominal price, alerts the poor on where to find drinking water; GroundTruth's Map Kibera, which enlisted residents of Nairobi's largest slum to map their neighborhoods and claim public services (Hagen, 2011); and Grameen Foundation's App Lab which allows poor Bangladeshi women to access and manage their finances (Taylor, Schroeder, & Meyer, 2014). Digital medical diagnostic software was launched in rural areas of the Himalayas where there is a dearth of doctors, providing an affordable service to the poor on diagnosis and treatment of medical problems (Arora, 2012).

While these are commendable efforts, we need to recognize that these are also business models that rest on the failings of the state. The longevity of such social entrepreneurship lies in the belief that the state will continue to disappoint its citizens. Here, zones of marginalization become zones of innovation. Besides, there are social values implicitly embedded in the design of these tools that influence the outcomes. Take, for instance, the farmers' app for agricultural guidance; it is clear that a number of assumptions reveal the naiveté of these well-meaning initiatives:

> [F]armers will not contest information that they receive online; . . . they will trust these new computing intermediaries and old intermediaries will disappear. It is assumed that farmers will be more receptive to agricultural information when receiving it online versus through other traditional communication media such as the radio and television. Further, it is believed that the dearth of relevant agricultural information is the prime reason holding farmers back from achieving mobility. Another presupposition is that given the access to such information, farmers will abandon their "traditional" practices for "better" agricultural practices, where efficiency and productivity are the prime goals. (Arora, 2010b, p. 127)

A case study on the implementation of medical diagnostic software in the Himalayas found that in surveying villagers on their health issues to populate the software, the programmed categories did not suffice. In fact, the majority of villagers reported illnesses that fell into the "Other" category, primarily due to social deprivations such as chronic hunger, long hours in the field, gender bias and practices like child marriage, and patronizing the local shaman (Arora, 2010a). Hence, to more effectively instrumentalize big data for the benefit of the poor, we need to incorporate these BoP users at the initial stages of software design and programming. This will enable sensitization of the socioeconomic and cultural contexts at hand and embed a more representative value system reflective of these new consumers. In tandem, we should

institute targeted literacy programs alongside the implementation of big data projects to educate designers and programmers on social practices that marginalize the most vulnerable segments in society (in this case, the lower castes and girls) if we are to operationalize the "inclusive" part of the inclusive capitalism model.

Bott and Young (2012) argue that crowdsourcing in the Global South comes with complex challenges. First, they question the nature of the "crowd" as necessarily active and diverse. The extent of activity depends on the degree of authoritarianism in the social fabric, allowing for participatory practices without endangering one's safety. The extent of "open" participation rests on the level of citizens' trust that their government will institute changes in policies and practices. In other words, just because databases are "open," they do not necessarily result in open practices. Second, regarding diversity, Bott and Young (2012) found that often the crowd does not emerge from the bottom, but rather from the top of the pyramid. Given the deep inequalities and often patriarchal structures in the Global South, it is not surprising that the authors found an overrepresentation of elite, educated, young males amid the digital crowd.

Third, much data that is valid for the poor is often nondigital, creating an obstacle to democratization through digital databased projects. For instance, the diagnostic and treatment app for the poor in the rural Himalayas was populated with only allopathic data (mainstream medicine practiced in the West), disregarding indigenous medical practices such as homeopathy, Ayurveda, and oral traditions (Arora, 2010a). Fourth, the poor have access to information via nonsmart mobile phones with limited and often expensive data plans (Rangaswamy & Cutrell, 2013), providing a reductionist and skewed consumption of shared intelligence. This can foster misunderstandings when readers consume only fragments of information, weakening rather than strengthening the crowd as informants. In such a digital climate the crowd is ripe for manipulation. Fifth, crowdsourcing allows the state to more easily target subjects, where contributors can be incriminated for their participation. This feeds into the rich literature on protests and digital media beyond the West, debunking the much-touted Twitter and Facebook revolutions (Kharroub & Bas, 2015). Crowdsourcing in the hands of the state can have deadly consequences by strengthening oppression of the masses, such as how digital technologies facilitated political violence by armed groups across Africa (Pierskalla & Hollenbach, 2013). Lastly, there is a false understanding that collaborative knowledge is synonymous with quality knowledge, given that these knowledge makers are not necessarily representative or experts (Arora & Vermeylen, 2013).

An independent project evaluation of Ushahidi, the poster child of crowdsourcing in the Global South, revealed that in spite of tremendous media attention, this program continues to face serious barriers in adoption and use (Morrow, Mock, Papendieck, & Kocmich, 2011). It was found that the data reporting often clashed with rigid information requirements of the international NGOs involved in disaster relief. Poor information infrastructures such as outdated computers and browsers and limited bandwidth contributed to low levels of use. Often, the messages from the field lacked sufficient detail, slowing relief-planning needs. There was a significant rate of misclassification, sometimes intentional, where volunteers would misclassify general distress messages for food or water to garner immediate attention from the relief organizations. Also, while the strength of Ushahidi lies in its field partnerships with other crisis

organizations, this created uncertainty in its corporate identity and, thereby, low awareness among the targeted user base in Haiti.

To conclude, social entrepreneurial efforts using big data in the Global South can extend democratic practice, but requires the involvement of BoP users at all steps, including software design and implementation. To keep us from romanticizing the values of the poor, we must ensure that literacy campaigns align with the implementation of these digital tools to guarantee they will be accessible to the most vulnerable populations among the BoP base. Further, while crowdsourcing for development can be instrumental in redistributing power over information and facilitating grassroots momentum, it can just as easily become a hegemonic mechanism of control. Indeed, the social impact of big data projects can have far more deadly consequences in the Global South, depending on the society's degree of patriarchy and authoritarianism. It is clear that crowdsourcing in the Global South has much overlap with the Global North on matters of concern such as garnering quality information, distributing expertise, addressing institutional politics in programming, and representing a diverse public. While the focus of this article is on big data in the Global South, we need to situate this neglected arena within the larger big data discourse to disrupt neat distinctions between the Global North and the Global South. In doing so, this opens up the field to a more nuanced analysis of how big data practices embody a constant interaction among global imaginations and governance models and local/national dynamics.

Databased Identity

In the section on databased democracy, the focus is on analyzing how big data is designed and implemented to foster an inclusive society by enhancing access to social systems. Here, identity is framed by the notion of citizenship. In this section we delve into how big data enables the identification, organization, and classification of groups and individuals, particularly those on the margins, and the consequences of such algorithmic structures. Clearly, both sections are related, and yet, identity does require going beyond the citizen to a more complex self.

The Postmodern Self, Politics of Algorithm and Cosmopolitanism from Below

Groups that are the most marginalized and vulnerable are often the most typecast, with group identities imposed on them. In the forgoing section, case studies underlined how the homeless, the migrants, the rural villagers, and women and girls become subservient to the politics of information infrastructures in their numerous attempts at databased legitimation. In this section, we reflect on how manufactured social narratives infuse databased structures and reinforce staid identities beyond the notion of citizenship. Numerous studies have been done on this subject such as the decades of institutionalizing black pathology as normative of African-American mental health (Williams & Williams-Morris, 2000), framing gender identities as "a cultural category and a material institution that uses biological differences to construct the sexual division of labour" (Castells, 2011, p. xxix), to the influential "culture of poverty" worldview, that argued that the values of poor communities perpetuated their state of poverty (Lewis, 1959). In the context of the Global South, the ground-breaking work by Narayan (2000) alerted development practitioners and scholars alike on the fact that few knew much about the target groups that they spent years researching and serving. She reported that the poor cared about much of the

same things as the rest of us: happiness, children, peace, dignity, safety, respect. Over the years rich and thoughtful critiques have emerged on framing the identities of the poor in these emerging economies. Sustained ethnographies have revealed the multidimensional nature of the poor, including their agency, their plurality and the "cosmopolitanism from below" (Appadurai, 2013).

New technologies have a way of promising social disruption in the form of allowances for multiple identities. In the early years of the Internet, there was much enthusiasm for virtual worlds of fantasy, where one could break free from social identities through plural selves of our own making:

> In the MUDS[2], the projections of self are engaged in a resolutely postmodern context. Authorship is not only displaced from a solitary voice, it is exploded. The self is not only decentered but multiplied without limit. There is an unparalleled opportunity to play with one's identity and to "'try out'" new ones. MUDS are a new environment for the construction and reconstruction of self. (Turkle, 1994, p. 158)

Since then, there has been a substantive critique of this so-called freedom of identity, arguing instead how social and political institutions, values, and rules of governance of the material world continue to affect the digital world (Dawson & Cowan, 2013; Gal, Shifman, & Kampf, 2015; Nakamura, 2013). Hence, we see racism, sexism, and other forms of power play continue online, with little escape from our enforced identities.

With every new technological innovation come new promises, new euphoria, and a regurgitation of past hopes and aspirations. Big data is no different. Technology becomes ahistorical—again. Jos de Mul (2015), a prominent philosopher of technology, celebrates the liberating aspect of big data, bringing back to life Turkle's postmodern self. To illustrate the power of big data in freeing up identity as opposed to the traditional database, he equates the relationship between big data and traditional databases to that of interculturalism and multiculturalism. This metaphorical argument underlines how the traditional database is much like multiculturalism, where different cultural identities (datasets) coexist, but do not interact with one another. In contrast, big data, much like interculturalism, is about intersectionality, interactivity, and play in infinite ways, revealing a more multiplexed identity. This allows the liberation of the postmodern self from her traditional codes. Big data speaks against stereotypes, allowing for what de Mul (2009 calls the "exhibition value" being replaced by "manipulation value" (p. 98). From digital reproduction, we are now in the age of digital recombination, a new database ontology. De Mul employs another metaphor to highlight big data's capacities: the gene versus the meme. Traditional data has been treated much like genes, permanent and unchanging, while big data allows data to transform into memes, dynamic, temporal, and noncommittal, with a possibility of infinite recombination. At last, de Mul proclaims, "In this digital era, tradition has become a commodity rather than an existential choice" (2009, p. 18). To illustrate this computational identity further, he brings up an example meant to underline the celebratory aspect of big data's "intercultural" dimension:

---

[2] Multi-User Dungeon or "MUD" is a multiplayer real-time virtual world.

> I saw an amusing example of this [interculturalism] last spring on a metro station in Rotterdam when a young, veiled Muslim girl came towards me: She was on rollerskates, wearing a Smiley T-shirt and talking, via her mobile phone, to a girlfriend in a remarkable mixture of Dutch and Arabic. (2015, p. 97)

Implicit here, however, is the pre-social sorting of these datasets, where, clearly, the dataset of "muslim girl" (religion and gender), "rollerskates" (sports), and "mobile phones" (technology) belong to different databases. The point of departure starts with sorting the datasets according to predisposed beliefs that view these categories as normatively separate. This privileged and problematic worldview is often programmed into the design of digital architectures, reproducing existing prejudices as a social fact. Hence, we need to pay greater heed to where the values in digital design emerge and who dictates these information infrastructures if we are to secure open-ended databased identities.

Stereotyped identities become overt on search engines, reenacting, reproducing, and reinforcing prevailing cultural codes. Take the autocomplete feature offered by search engines such as Google. Algorithms shaped by users' aggregated search activities on the Web dictate every search query. Far from providing us with infinite combinations and new associations with identities, it was found that search queries underlined and exacerbated racism, sexism, and political difference (König & Rasch, 2014). Hence, it is impossible to separate online and offline values, social practices, and power relations as they mutually reconstitute each another.

Anish (2015) extends this discourse with big data developments in the Global South by problematizing how datafication is being framed as a way to replace "social identity" (e.g., face-to-face conversations, village-based kinship, place-based kinship) with "system identity":

> While social identity is an identity continually renegotiated through linguistic interactions and social performances, bureaucratic identity—glimpsed in passports, driver's licenses, and other identity cards—is a construction of fixed personhood for the purposes of modern organizational needs, ensuring that the member has remained essentially the same despite changes in personality, body, and behavior. With the spread of information technologies, however, there has emerged a new variation of identity—system identity, which represents persons as dynamically forming clouds of data. While system identities can serve the bureaucratic need for identifying members, their role far surpasses the functional necessities of inclusion and exclusion. (p. 42)

There is the perception that as you scale up, you trade identity for inclusion. The reductionist view of village life and even disdain toward the poor are reminiscent of the culture-of-poverty worldview that often become inscribed in the design of big data architectures. The postmodern self is a fiction and fantasy that resurfaces with each new technological innovation. It is equated with a democratic state of being. While this argument has played out numerous times in the past, big data seems to have inspired an encore. Features such as autocomplete reproduce stereotypes by reinforcing negative associations. Clearly, social identities are curated online as commercial interests, state agendas, and powerful cultural groups exercise their agency in this identity game. Unlike the Global North where policies and laws are

being put in place to protect individuals and organizations from damaging associations (e.g., the right to be forgotten), much of the Global South has yet to prioritize the regulation of big data to protect its citizens from harmful digital practices. Hence, transparent and open data curation of group and individual identities can lead to more thoughtful policies and practices in these emerging digital economies.

Databased Geography

Prior sections bring to the surface deep-seated biases on how big data is framed, designed, and instrumentalized in the regions of the Global South. This article argues that big data projects in these marginalized contexts are predominantly driven by the rhetoric of democracy, inclusion, and empowerment of the poor, subjecting these seemingly altruistic efforts to far less scrutiny than they deserve. By offering a conceptualization of databased democracy and databased identity, the intent here is to critique the extent to which traditional social practice is redefined and transformed with the rise of big data. However, it is not sufficient to look at the process of change that big data may or may not engender, but also the nature and digital locatedness that these enactments emerge from and occupy. Hence, this section on databased geography argues that decades of development projects in the Global South have led us to analyze BoP users' digital behavior as utilitarian and, thereby, not representative of this rich and diverse public.

Leisure Commons and the Playful Poor in Emerging Economies

It can be argued that the digital commons is today's leisure commons (Arora, 2014). At this stage, we have much evidence that the most frequented sites in the Global North as well as the Global South are leisure oriented. Primarily, people go online to romance, game, be entertained, consume media, view pornography, and share their personal thoughts and feelings. It is the arrival of a new kind of movement, a novel way to experience, produce, and consume leisure: "whether desired or not as part of any "official" history of this currently central cultural medium, online recreation or "virtual leisure" has been positioned among the dominant elements within the Internet's development" (Weiss, Nolan, Hunsinger, & Trifonas, 2006, p. 961). Hence, digital geographies such as Facebook, Twitter, YouTube, and other networking sites are prime social geographies within which people and organizations interact, protest, and commercialize their engagements. There is tremendous research being done on how big data within these terrains is harnessed for multiple purposes from the personal, the commercial, and the political, immersing ourselves in debates on privacy, surveillance, digital labor, and prosumption (Postigo, 2014).

When we shift our attention to the Global South and the BoP users, their new media practices are predominantly framed as instrumental and utilitarian. This is partly because development agendas drive this research with a strong historical bias for the socioeconomic (Arora & Rangaswamy, 2013). For instance, there is more emphasis on farmers checking crop prices online than, for instance, watching pornography on their mobile devices. Growing scholarship, however, underlines digital leisure practices among the poor in recent years: how slum youths use mobile phones to romance girls in India (Rangaswamy & Cutrell, 2013), play games in LAN houses in Brazil (Kolko & Racardio, 2014), access entertainment via mobile banking apps in Kenya (Gajjala & Tetteh, 2014), and generate political jokes in

the Twittersphere and blogosphere in China (Yang & Jiang, 2015). These new media cultures underline the importance of desire as an aspirational construct, making visible the diversity of this emerging public (Udupa, 2015).

Clearly, a mental shift is needed for how to approach this populace and their user practices if we are to gain a nuanced and comprehensive understanding of the social impact of big data in their daily lives. This article argues that to understand the nature of big data in the Global South, we need to first recognize that digital life is lived within digital leisure geographies for the most part. As emerging economies globalize and urbanize exponentially and their BoP users, or "prosumers," become more critical consumers and creative contributors of digital content and, arguably, free laborers instead of classic development beneficiaries, this paradigm shift will provide an open-ended, explorative, and pluralistic perspective on big data research.

Conclusion

There is no doubt that the data deluge produced by the bottom-of-the-data-pyramid users will have a major impact on the future of the Internet. The question remains - How do we treat this rising populace as culturally diverse and yet refrain from exoticizing them? We must be careful not to romanticize the poor, as they are a pluralistic group that at times perpetuates social values harmful to certain segments of society based on, for instance, caste, tribal affiliation, and gender. Digital literacy campaigns should accompany big data projects in the design and implementation, negotiating human rights with culturally sensitive practices. While we give due recognition to big data as an empowering tool within emerging economies, we must attend to simultaneous efforts to strengthen institutions that will protect individual and group privacy. In this pursuit, alternative modes of inclusivity should be sought beyond the default neoliberal approach on the marketization of the poor through inclusive capitalism.

Current studies on the "big data divide" (Andrejevic, 2014; Boyd & Crawford, 2012) make peripheral the dichotomies of the West and developing nations to the more central focus on systemic inequality between those who have access to and control over data and those who do not. This article argues that while these studies tackle big data through the lens of power relations on a globalized stage, we need concerted and sustained scholarship on the role and impact of big data on the Global South. After all, the technocratic approach of the state through ICT-oriented development projects has a distinct colonial legacy in much of the Global South. Currently, user behavior and institutional practices of the Global North disproportionately represent and influence our understandings on this matter, which can serve as a genuine barrier to thoughtful, indigenous design of big data applications for emerging economies.

The coupling of the term "databased" with democracy, identity, and geography is deliberate as it allows us to focus on expectations, assumptions, and prevalent policies and practices, instigated by the rise of big data. Here, big data promises to foster a new type of inclusion, personhood, and sense of place. This article analyzes this essentializing of technology and joins a rich scholarship on the problematizing of utopian claims of past mass communication technologies, including the radio, television, and computer. Unfortunately, as this article discusses, these discourses are persistent and appealing and have gained a

new lease on life in the age of big data. This work illustrates the challenges of reifying this idyllic notion by questioning the premise that these surveillance architectures empower the poor. Anonymity has served as a powerful instrument of activism in these suppressed contexts and continues to this day as witnessed in the uprisings in the Middle East and elsewhere. Thereby, the pretext of associating anonymity with vulnerability negates larger understandings and histories of identification as a technology of control. Hence, we need to disrupt the innate belief in the harmony of the state with that of data of the citizen.

While the governmentality of big data can produce major efficiencies for the poor such as the disseminating of welfare benefits through the biometric identity project, it comes at the astounding price of privacy. While several studies today indicate that people in the West attribute high value to privacy, there is a dearth of studies on how marginalized populations in the Global South view, construct, and practice privacy. Thereby, the trade-off is made on their behalf, with little involvement of this substantial public. While several countries such as The Netherlands, United Kingdom, and France are declaring data consolidation unconstitutional in order to protect their citizens, in the Global South this trend is moving in the opposite direction. Wide-sweeping and wretched poverty provide the urgency and excuse to distance events and policies of the Global South from those of the Global North. Guilt politics and moral blackmail interfere in the democratic shaping of these systems, building on the rich and global learnings of instituting digital mechanisms for equitable and fair practice. Hence, the dichotomy of the Global North–South, while artificial, serves to identify and dismantle exceptionalism and exoticism prevalent in the big data efforts in the Global South.

When it comes to BoP economies of doing good and making a profit, marketizing the poor is positioned as an innovative effort to include the marginalized as consumers and agents of change. Again, trade-offs are made; through Facebook's Free Basics platform, net neutrality is sacrificed to give the poor free access to certain Internet sites. Is net neutrality a privilege for only consumers in the Global North? These short-term measures fortify Facebook's dominance in the digital market of the Global South, delivering exclusive insights into the databased behavior of the BoP demographic. However, it is common knowledge that monopolies rarely operate for the common good, often reversing initial benefits as power becomes concentrated in select entities.

Our point of departure on today's information economy should be empirically driven instead of ideological. We have amassed much evidence to date on how racism, sexism, and other forms of power play continue online, reinforcing stereotypes. The Internet has not and will not naturally manifest in a virtual community of self-organized equality as envisioned in the early years. Clearly, social identities are curated online as commercial interests, state agendas, and powerful cultural groups exercise their agency in this identity game. As the residents of the Global South move exponentially online, particularly the disenfranchised, they intersect with local and global politics and economies. Their digital activity requires much-needed scrutiny to gain a globalized worldview on big data. We need to provide institutional, financial, and social encouragement to grassroots activism, where online representation of the currently invisible and vulnerable is brought to the fore. The "cosmopolitanism from below" (Appadurai, 2013) should find its way to the big data stage. We must pay more attention to where the values in digital design emerge and who dictates these information infrastructures to create allowances for a richer databased identity. The Global South should be actively engaged with current debates—such as the right

to be forgotten—as multinational IT companies confront national sentiments, values, and institutions, illustrating how context continues to matter.

Lastly, we must recognize the cultural character of the digital sphere in which the poor live and act. Amassed evidence over the decades reveals that what users primarily do online is characterized by entertainment, romance, gaming, and socializing. We need to insert the notions of leisure and desire into our analytical framework when approaching the social impact of big data produced and consumed by this rising public. Databased geographies are leisure geographies, even for the poor. This opens fruitful avenues of scholarship, bridging leisure studies with new media studies, which is not the prerogative only of the Global North. Yet, most studies of new media practices among the poor in emerging economies continue to focus on the instrumental aspects. Hence, a critical reconsideration of embedding the aspirations, desires, values, and behaviors of this largely unexamined populace into big data infrastructures is the pathway to a democratic digital sphere.

References

Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, *8*, 1673–1689.

Anish, A. (2015). *Neutral accent: How language, labor, and life become global*. Durham, NC: Duke
    University Press.

Appadurai, A. (2013). *The future as cultural fact: Essays on the global condition*. London, UK: Verso.

Arora, P. (2010a). Digital gods: The making of a medical fact for rural diagnostic software. *The
    Information Society*, *26*(1), 70–79.

Arora, P. (2010b). *Dot com mantra: Social computing in the Central Himalayas*. Oxford, UK: Ashgate
    Publishing.

Arora, P. (2012). Leisure divide: Can the third-world come out to play? *Information Development, 28*(2),
    93–101.

Arora, P. (2014). *Leisure commons: A spatial history of Web 2.0*. Oxford, UK: Routledge, Taylor &
    Francis.

Arora, P., & Rangaswamy, N. (2013). Digital leisure for development: Rethinking new media practices
    from the Global South. *Media Culture & Society*, *35*(7), 898– 905.

Arora, P., & Vermeylen, F. (2013). The end of the art connoisseur? Experts and knowledge production in
    the visual arts in the digital age. *Information, Communication & Society*, *16*(2), 194–216.

Blowfield, M., & Dolan, C. (2014). Bottom billion capitalism: The possibility and improbability of business as a development actor. *Third World Quarterly*, *35*(1), 22–42.

Bott, M., & Young, G. (2012). The role of crowdsourcing for better governance in international development. *PRAXIS The Fletcher Journal of Human Security, XXVII,* 47–70.

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679.

Breckenridge, K. (2005). The biometric state: The promise and peril of digital government in the new South Africa. *Journal of Southern African Studies*, *31*(2), 267–282.

Castells, M. (2011). *The power of identity: The information age: Economy, society, and culture* (Vol. 2). New York, NY: John Wiley & Sons.

Couldry, N., & Powell, A. (2014). Big data from the bottom up. *Big data and society*, 1–5. doi:10.1177/2053951714539277

Dawson, L. L., & Cowan, D. E. (2013). *Religion online: Finding faith on the Internet*. New York, NY: Routledge.

De Mul, J. (2009). The work of art in the age of digital recombination. In M. v. d. Boomen, S. Lammes, A.-S. Lehmann, & M. T. Schäfer (Eds.), *Digital material: Anchoring new media in daily life and technology*. Amsterdam, The Netherlands: Amsterdam University Press.

De Mul, J. (2015). Database identity: Personal and cultural identity in the age of global datafication. In W. de Been., P. Arora, & H. Hildebrandt (Eds.), *Crossroads in new media, identity and law: The shape of diversity to come*. London, UK: Palgrave.

Dolan, C., & Roll, K. (2013). Capital's new frontier: From "unusable" economies to bottom-of-the-pyramid markets in Africa. *African Studies Review*, *56*(3), 123–146.

Eagle, N. (2015, June 1). How to make the Internet free in developing countries. Crunch Network. Retrieved from http://techcrunch.com/2015/06/01/how-to-make-the-internet-truly-free-in-developing-countries

Gajjala, R., & Tetteh, D. (2014). Relax, you've got M-PESA: Leisure as empowerment [Special Issue]. *Information Technologies & International Development, 10*(3), 31–46.

Gal, N., Shifman, L., & Kampf, Z. (2015). "It gets better": Internet memes and the construction of collective identity. *New Media & Society*, 1–17. doi:10.1177/1461444814568784

GSMA. (2014). *GSMA works with the ministry of communication and information technology to support development of the mobile sector in Myanmar*. Retrieved from http://www.gsma.com/newsroom/press-release/gsma-ministry-communications-myanmar/

Hagen, E. (2011). Mapping change: Community information empowerment in Kibera (Innovations Case Narrative: Map Kibera). *Innovations*, *6*(1), 69–94.

Hamel, G., & Prahalad, C. K. (2013). *Competing for the future*. Cambridge, MA: Harvard University Press.

Hellström, J. (2015). Crowdsourcing as a tool for political participation? The case of Uganda Watch. *International Journal of Public Information Systems*, *11*(1), 1–19.

Hilbert, M. (2013). *Big data for development: From information to knowledge societies*. Retrieved from http://ssrn.com/abstract=2205145

Kharroub, T., & Bas, O. (2015). Social media and protests: An examination of Twitter images of the 2011 Egyptian revolution. *New Media & Society*, 1–20. doi:10.1177/1461444815571914

Kolko, B., & Racadio, R. (2014). The value of non-instrumental computer use: A study of skills acquisition and performance in Brazil [Special Issue]. *Information Technologies & International Development*, *10*(3), 47–65.

König, R., & Rasch, M. (Eds.). (2014). S*ociety of the query reader. Reflections on Web search.* Amsterdam, The Netherlands: Institute of Network Cultures.

Lewis, O. (1959). The culture of poverty. *Scientific American*, *4*(215), 19–25.

Linders, D. (2013). Towards open development: Leveraging open data to improve the planning and coordination of international aid. *Government Information Quarterly, 30*(4), 426–434.

Lyon, D. (2013). *The electronic eye: The rise of surveillance society—Computers and social control in context.* New York, NY: John Wiley & Sons.

Magnet, S. A. (2011). W*hen biometrics fail: Gender, race, and the technology of identity*. Durham, NC: Duke University Press.

Morrow, N., Mock, N., Papendieck, A., & Kocmich, N. (2011). Independent evaluation of the Ushahidi Haiti project. Retrieved from https://www.ushahidi.com/blog/2011/04/19/ushahidi-haiti-project-evaluation-final-report

Nakamura, L. (2013). *Cybertypes: Race, ethnicity, and identity on the Internet*. New York, NY: Routledge.

Narayan, D. (2000). *Voices of the poor: Can anyone hear us?* Oxford, UK: Oxford University Press.

Owen, T. (2014). Foucaultian dispositifs as methodology: The case of anonymous exclusions by unique identification in India. *International Political Sociology*, *8*(2), 164–181.

Parker, I. (2011, October 3). The I.D. man: Can a software mogul's epic project help India's poor? *The New Yorker*. Retrieved from http://www.newyorker.com/magazine/2011/10/03/the-i-d-man

Philip, K., Irani, L., & Dourish, P. (2012). Postcolonial computing: A tactical survey. *Science, Technology, & Human Values*, *37*(1) 3–29.

Pierskalla, J., & Hollenbach, F. (2013). Technology and collective action: The effect of cell phone coverage on political violence in Africa. *American Political Science Review*, *107*(2), 207–224.

Postigo, H. (2014). The socio-technical architecture of digital labor: Converting play into YouTube money*. New Media & Society*, *3*(4), 17–24.

Pötzsch, H. (2015). The emergence of iBorder: Bordering bodies, networks, and machines. *Environment and Planning D: Society and Space*, *33*, 101–118. Retrieved from http://epd.sagepub.com/content/33/1/101.abstrac

Prahalad, C. K. (2009). *The fortune at the bottom of the pyramid, Eradicating poverty through profits* (Rev. ed.). New Jersey, NJ: FT Press.

Rai, S. (2012, June 6). *Why India's identity scheme is groundbreaking.* Retrieved from http://www.bbc.com/news/world-asia-india-18156858

Rangaswamy, N., & Cutrell, E. (2013). Anthropology, development and ICTs: Slums, youth, and the mobile Internet in urban India. *Information Technologies & International Development, 9*(2), 51–63.

Rao, U. (2013). Biometric marginality UID and the shaping of homeless identities in the city. *Economic & Political Weekly*, *48*, 1–7.

Roy, A. (2012). Ethical subjects: Market rule in an age of poverty. *Public Culture*, *24*(1), 105–108.

Sarkar, S. (2014). The unique identity (UID) project, biometrics and re-imagining governance in India. *Oxford Development Studies*, *44*(4), 516–533.

Sengoopta, C. (2003). *Imprint of the raj: How fingerprinting was born in colonial India*. London, UK: Macmillan.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York, NY: Little Brown.

Taylor, L., & Schroeder, R. (2014). Is bigger better? The emergence of big data as a tool for international development policy. *GeoJournal, 80*(4), 503–518.

Taylor, L., Schroeder, R., & Meyer, E. (2014). Emerging practices and perspectives on big data analysis in economics: Bigger and better or more of the same? *Big Data & Society*, *1*(2), 1–10.

Toness, R. V. (2014, December 10). *India building database to unite records for 1.2 billion*. Retrieved from http://www.bloomberg.com/news/articles/2014-12-10/india-building-database-to-unite-records-for-12-billion

Turkle, S. (1994). Constructions and reconstructions of self in virtual reality: Playing in the MUDS., *Mind, Culture, and Activity*, *1*(3), 158–167.

Udupa, S. (2015). *Making news in global India: Media, publics, politics*. Cambridge, UK: Cambridge University Press.

Van der Ploeg, I. (2005). *The machine readable body: Essays on biometrics and the informatization of the body*. Maastricht, The Netherlands: Shaker.

Weiss, J., Nolan, J., Hunsinger, J., & Trifonas, P. (Eds.). (2006). *International handbook of virtual learning environments*. New York, NY: Springer.

Williams, D. R., & Williams-Morris, R. (2000). Racism and mental health: The African-American experience. *Ethnicity & Health*, *5*(4), 243–268.

Yang, G., & Jiang, M. (2015). The networked practice of online political satire in China: Between ritual and resistance. *International Communication Gazette*, *77*(3), 215–231.