Title
Using Crowd-Sourced Data to Study Public Services: Lessons from Urban India

Journal
Studies in Comparative International Development, 53(3)

ISSN
0039-3606

Authors
Post, Alison E
Agnihotri, Anustubh
Hyun, Christopher

Publication Date
2018-09-01

DOI
10.1007/s12116-018-9271-4

Peer reviewed

Using Crowd-Sourced Data to Study Public Services: Lessons from Urban India

Alison E. Post
University of California, Berkeley
aepost@berkeley.edu

Anustubh Agnihotri
University of California, Berkeley

Christopher Hyun
University of California, Berkeley

Forthcoming, Studies in Comparative International Development

As cities throughout the developing world grow, they often expand more quickly than the infrastructure and service delivery networks that provide residents with basic necessities such as water and public safety. Why do some cities deliver more effective infrastructure and services in the face of rapid growth than others? Why do some households and communities secure better services than others? Answering these questions requires studying the large, politicized bureaucracies charged with providing urban services, and especially the relationships between frontline workers, agency managers, and citizens in informal settlements. Researchers investigating public service delivery in cities of the Global South, however, have faced acute data scarcity when addressing these themes. The recent emergence of crowd-sourced data offers researchers new means of addressing such questions. In this paper, we draw on our own research on the politics of urban water delivery in India to highlight new types of analysis that are possible using crowd-sourced data, and propose solutions to common pitfalls associated with analyzing it. These insights should be of use for researchers working on a broad range of topics in comparative politics where crowd-sourced data could provide leverage, such as protest politics, conflict processes, public opinion, and law and order.

WORD COUNT: 8696

Acknowledgments:
This research was funded by a "DIL Innovate" Grant from the Development Impact Laboratory, Blum Center for Developing Economies (USAID Cooperative Agreement AID-OAA-A-13- 00002, Alison Post and Isha Ray Principal Investigators), U.C. Berkeley, and a dissertation fieldwork grant from the Institute for International Studies, U.C. Berkeley. Tanu Kumar and Isha Ray, U.C. Berkeley, are co-authors of the impact evaluation project described in this paper. We thank Maria Chang for research assistance. We also thank NextDrop and the Bangalore Water Supply and Sewerage Board (BWSSB) for their support of our research. We are grateful for comments from Thad Dunning, Agustina Giraudy, Tanu Kumar, Katerina Linos, Aila Matanock, Isha Ray, and

seminar participants at U.C. Berkeley and American University.

As cities throughout the developing world grow, they often expand more quickly than the infrastructure and service delivery networks that provide residents with basic necessities such as clean water, sanitation, electricity, trash collection, and public safety. Even when states do expand the territorial reach of services, they often fail to maintain existing infrastructure, detracting from service quality.

Why do some cities offer more effective infrastructure and services in the face of rapid urban growth than others? Why do some households and communities secure better services than others? Answering these questions requires studying the large, politicized bureaucracies charged with providing urban services. Urban service providers tend to be much larger, more complex entities than the organizations tasked with delivering comparable services in rural settings. Their complexity stems, on the one hand, from the more complicated nature of the infrastructure required to service urban populations. Whereas rural and suburban areas typically rely on well water, high-density urban centers require large-scale, technically complex treatment and distribution systems, parts of which must be managed by specialists. Similarly, rural road and rail networks are not as technically complex as urban subways, bus rapid transit systems, or ring roads and associated pedestrian overpasses. The intricacy of urban service bureaucracies also stems from the greater number of personnel required to cater to the large and diverse populations receiving services. As a result, urban service delivery bodies tend to include elaborate organizational hierarchies with specialized roles, creating not only coordination difficulties, but also significant informational asymmetries between employees and supervisors.[1]

Investigating variation in the quality and allocation of urban services requires examining decision-making within these extensive organizational hierarchies and paying close attention to interactions between frontline workers and citizens. It is particularly important to examine these bureaucracies' activities and relationships with citizens in informal settlements. Slums are, after all, where the bulk of the low-income population lives in cities of the Global South. Slum populations possess great need for basic infrastructure services such as water and policing, because private substitutes—where they exist—are often much more expensive and of poorer quality. Examining decision-making by different actors in these organizations, as well as their relationships with citizens, can shed light on why urban service providers extend state services, or provide better services, in some settlements than in others. Such analysis can also help one understand patterns of indirect provision in which public employees or elected officials allow private entrepreneurs or other actors to unofficially mediate access to state services. For example, private water tankers may purchase bulk water from utilities (illegally or legally); similarly, independent electricity distributors may sell power siphoned off from state networks while state linemen look the other way (see Post, Bronsoler, & Salman, 2017).

Considering the complex and highly politicized nature of service provision in informal settlements prompts us to revisit important theoretical debates in comparative

[1] As Scott (1996, pp. 74–75) famously noted, such information asymmetries can arise from experiential knowledge ("mētis") as well as technical skills.

politics. In particular, it encourages us to examine street-level bureaucrats (SLBs) alongside more commonly studied intermediaries such as partisan brokers and neighborhood leaders, and pose questions that have received little attention within political science in recent decades. To what extent do SLBs exercise significant discretion—vis-à-vis their superiors because of information asymmetries in these complex bureaucracies—and make policy on the ground, as argued by Lipsky (1980)? When they do possess significant autonomy, how do they choose to allocate services between groups or households? Do they respond to incentives similar to or different from those faced by intermediaries more commonly described in the literature? To what extent do SLBs balance the demands of "multiple principals," including not just their organizational superiors, but also local politicians, neighborhood leaders, and citizens?

One of the reasons we know so little about the politics of urban service provision in the Global South—and especially about the direct interface between citizens and providers—is that data is scarce due to information asymmetries and a lack of systematic data collection. While these problems plague all large bureaucracies, they take on stronger forms in cities of the developing world due to lower levels of state capacity and electronic record-keeping. The heads of service delivery bureaucracies often know very little about when and where water is flowing through a network, where beat cops are patrolling at a given point in time, or which teachers have arrived at school. They often possess little systematic data related to how citizens experience or perceive service delivery. It is therefore difficult to disentangle service delivery failures due to political will or principal-agent problems from those stemming from simple lack of data needed for oversight.

The recent emergence of "crowd-sourced data," however, offers researchers new means of addressing such questions, as well as a range of other topics in comparative politics, such as protest dynamics, political conflict, public opinion, and corruption. *Crowd-sourced data* refers to information sourced electronically from disparate individuals or groups. With respect to urban public services, data sourced from citizens (the "crowd" as typically conceived) can shed light on the quality and reach of services, as well as the performance of street-level bureaucrats. Data sourced from SLBs can also describe service quality and access, as well as SLB activities themselves.

In this paper, we draw on our own research on the politics of urban water delivery in India to highlight new types of analyses that are possible using crowd-sourced data, and propose solutions to common pitfalls associated with analyzing it. Specifically, we work with data collected by a social enterprise, NextDrop, to ease the burden imposed by intermittent water supply through crowd-sourcing. NextDrop incentivized water valvemen, the water utility employees charged with turning water off and on for particular city districts, to notify the company when they are opening valves via an automated voice response system. NextDrop used this data to send text message notifications to households regarding when they should expect their water to arrive. Households, in turn, were asked to verify the accuracy of the reports that they receive through the same electronic system.

NextDrop's novel data positioned us to speak to several important questions regarding the politics of urban service delivery and, especially, regarding the activities of SLBs. Our analysis suggests that the political economy literature on local public goods provision should pay far greater attention to SLB discretion and how decisions made by

SLBs affect service allocations, and service quality, and citizen perceptions of the state. Our work also suggests that political science scholarship on intermediaries such as partisan brokers and neighborhood leaders should pay far greater attention to how individual characteristics such as familial and financial circumstances affect their behavior.

Our analysis also shows that while crowd-sourced data offers an important supplement to absent or biased sources of information, it must be treated with care. We show how inferential challenges such as selective reporting, inaccurate reports, information-sharing, and technical and organizational challenges present important hurdles for researchers. We then illustrate approaches to addressing these inferential and organizational challenges through groundtruthing, researching the selection process through which data is contributed, employing cluster- rather than household-level randomization to minimize information-sharing, and research partnerships.

THE POTENTIAL OF CROWD-SOURCED DATA

"Crowd-sourcing" has risen in prominence with increases in mobile phone and internet penetration, as well as growing digital literacy, worldwide. Drawing on a recent meta-analysis (Estellés-Arolas & González-Ladrón-de-Guevara, 2012), we define *crowd-sourcing* as problem-solving or knowledge creation through distributed, electronic information-gathering processes. In other words, organizations make use of the expertise and efforts contributed voluntarily by disparate members of a "crowd." This crowd can vary in size and characteristics, depending on the nature of the desired task (Estellés-Arolas & González-Ladrón-de-Guevara, 2012, p. 6). Contributors of information can reside inside or outside of the organization initiating the crowd-sourcing processes, which in some cases includes only a limited number of employees (Estellés-Arolas & González-Ladrón-de-Guevara, 2012, p. 6). Familiar examples include Wikipedia, Amazon's Mechanical Turk, Waze, and the use of social media such as Facebook to mobilize collective efforts. We focus here specifically on *crowd-sourced data*, that is data or knowledge collected through crowd-sourcing processes. People serve as "sensors" or repositories of knowledge, which organizations can tap through new telecommunications technologies.

While crowd-sourced data is most prevalent in OECD countries given their high rate of smartphone penetration, it is increasingly available in the Global South, especially in cities. Basic cellphones, which can support text message-based crowd-sourcing, are now extremely common in cities of the developing world. In India, where our study is based, there are more than a billion mobile phone users with close to 603 million in urban areas. Access rates are very high in cities, with 165 mobile connections per 100 residents in urban areas, compared with 53 per 100 residents in rural areas (Telecom Regulatory Authority of India, 2017, p. i). Residents of informal urban settlements are typically equipped with cell phones. Meanwhile, smartphones—which allow for more complex forms of crowd-sourcing—are increasingly common. As prices have fallen, usage across the developing world has increased dramatically: smartphone ownership rates rose from a median of 21% in 2013 to 37% in 2015 (Poushter, 2016). The advent of cheaper smartphone technology suggests that crowd-sourced data will be increasingly useful for the study of rural phenomena as well. For example, recent low-intensity riots in small, rural towns in India were reportedly instigated using the messaging app, WhatsApp (Raza, 2017; Yadav, 2015).

Researchers, policymakers, activists, and citizens have begun to turn to crowd-sourced data as a means of understanding social and political phenomena, especially where state-collected data is scarce. Non-governmental organizations (NGOs), governments, businesses, and—to a lesser extent—academics are pioneering the use of crowd-sourced data to better understand the quality and reach of local public services, and improve their delivery.[2] Some systems solicit information from citizens regarding the quality, timing, and reach of services, as well as service gaps, integrating this information into real time information systems. For example, the Water Integrity Network, an NGO working in India, launched an app that crowd-sourced information about water, air, and soil quality (Sachdev, 2017). Similarly, the Mtrac application allows UNICEF to monitor disease outbreaks based on real time reports from health workers (UNICEF, 2012b). City-administered call centers and web-based complaint systems, such as New York's 311 system and Boston's "CitizensConnect" app, amass new information from citizens on local service deficiencies such as potholes and burnt out street lights. The transportation sector has perhaps witnessed the most innovation of this type (Offenhuber, 2017). Waze, for example, collects information from users of its mobile phone app regarding traffic conditions. Academics have also crowd-sourced data from riders to map paratransit routes and provide information on traffic conditions in developing countries (Ching, Zegras, Kennedy, & Mamun, 2013; Klopp et al., 2014).

[2] See World Bank (2016, Chapter 3) for a more extensive catalog of such initiatives.

Governments, NGOs, and researchers have also used crowd-sourced data to obtain better information about policy areas in which citizens tend to underreport problems through official channels. State data regarding crime, corruption, and service deficiencies, for instance, can be wildly inaccurate in countries with low state capacity because citizens often expect reports will not be acted upon, and may even fear the consequences of reporting. A variety of initiatives have encouraged citizens to report problems anonymously online. For example, the Brazilian WikiCrimes project validates crowd-sourced crime reports and shares them with local officials (Furtado, Caminha, Ayres, & Santos, 2012). In India, I Paid a Bribe crowd-sources information on corruption, and in Tanzania, the Taarifa web-application encourages communities to report on their problems related to water and sanitation (Iliffe, Sollazzo, Morley, & Houghton, 2014).

Governments and aid organizations increasingly crowd-source citizen feedback on spending priorities and other aspects of public policy. For example, local governments in India and Brazil have launched electronic systems to solicit citizen input on budgetary priorities (Touchton & Wampler, 2014).[3] UNICEF has launched U-report in Uganda, an application that allows citizens to voice their opinions on particular development projects (UNICEF, 2012a). Research on the effectiveness of such citizen engagement is in its infancy; more research is required to understand the linkages between crowd-sourced technology, citizen engagement, and policy outcomes (Peixoto & Fox, 2015).

[3] See also: http://timesofindia.indiatimes.com/city/delhi/Delhi-budget-to-be-crowdsourced-Arvind-Kejriwal/articleshow/46362366.cms

While we focus here on applications to public service delivery, researchers and policymakers have also begun using crowd-sourced data to examine a range of other phenomena of interest to political scientists. Some efforts focus on the measurement of political attitudes or ideology (Barberá, 2015; Boutet, Antoine, Hyoungshick Kim, and Eiko Yoneki, 2012; Calvo, 2015; Jamal, Keohane, Romney, & Tingley, 2015), while others have used crowd-sourced data to map political protests and understand their dynamics (Barberá & Metzger, 2013; Breuer, Landman, & Farquhar, 2012; Starbird & Palen, 2012; Vaccari et al., 2015). For example, researchers can conduct text analysis to analyze retweets and comments, aggregate data to estimate protest intensity and duration, and conduct social network analysis to find patterns (e.g., Barberá, 2013; Barberá & Metzger, 2013). Scholars have also tracked conflict activity using similar techniques (e.g., van der Windt & Humphreys, 2014). Others have used crowd-sourced data from social media and other sources to measure and understand undocumented behaviors, such as movements by unregistered refugees (Carlson, Jakli, & Linos, Forthcoming); similar data could also be used to understand the activities of informal sector workers. These potential applications of crowd-sourced data to political phenomena are summarized in Table 1.

Table 1. Existing and Potential Applications for Crowd-Sourced Data

Phenomenon	Measurement Application
Service delivery	• Reach and allocation of services • Service quality, deficiencies • Principal-agent relationships in bureaucracy • Frontline worker interactions with citizens (including corruption) • Citizen perceptions, preferences, priorities
Protest and conflict politics	• Intensity, duration, location of protests or conflict • Mobilization strategies • Characterizing participant networks

Public opinion	• Alternative source of data on political attitudes and ideology • Relationship between attitudes and exposure to networks, media, etc.
Law and order	• Crime rates • Corruption among police • Behavior of undocumented populations or informal sector activity

NEXTDROP'S CROWD-SOURCED DATA ON WATER INTERMITTENCY

NextDrop's crowd-sourcing system, the source leveraged in our research, was developed to collect and disseminate data on a critical dimension of service quality in urban public services: service frequency and predictability. Service frequency and predictability has received very little attention in the literature, in large part because so little data on it exists. Throughout the developing world, intermittency and unpredictability are the hallmarks of public service delivery: buses do not run on a standard schedule, water supplies are variable in terms of arrival times, and electricity blackouts occur unexpectedly. For example, 400 million people worldwide rely on intermittent water, often receiving services only a few days a week for a few hours (van den Berg & Danilenko, 2011). Published water schedules in newspapers or on the walls of local water offices often depart significantly from actual practice because utilities themselves lack the resources to purchase the sensors necessary to produce real time information regarding water allocations.[4] Service unpredictability can be particularly onerous for low-income populations, because someone must spend time waiting at home for water arrival

[4] While water meters can tabulate flow through specific pipes and connections, these are typically read manually at regular intervals, and thus do not give utilities real time information on how often and when water is delivered to particular areas.

to be able to fill household storage containers. Substitutes such as vended water tend to be much more expensive than municipal water. Higher income households, in contrast, can afford pumps that automatically fill household tanks when water services commence, as well as the load-bearing roofs that such tanks require.

NextDrop developed a novel, crowd-sourced approach to generating information about service timing, intended to help households cope with water intermittency. In their system, utility employees called “valvemen” called a toll-free number when opening and closing valves in specific “valve areas.” NextDrop then sent free text message notifications to individual households, which it had cataloged by valve area through the collection of household GPS coordinates, to let them know when their water would arrive.[5] NextDrop then “pinged” households to ask if delivery notifications were in fact accurate and provided households with a cost-free way to respond. To correctly place households in valve areas, NextDrop created valve area maps, which many Indian utilities do not possess. It did so by drawing on the valvemen’s tacit knowledge regarding the area boundaries, accompanying them on walks around the edges, and taking GPS readings. (Each polygon in Figure 1 is an example valve area from Bangalore: the city has thousands of such valve areas, which are areas of 50-200 households for which valvemen turn water off and on by manually adjusting a valve.) NextDrop then assembled valvemen report data suggesting where water was flowing throughout the utilities’ network into a “dashboard” that they shared with the utility’s engineers.

Figure 1. Example valve areas in Bangalore, India

[5] NextDrop’s revenue model involved charging utilities for information services, including real time information of water flows.

Valve areas from Subdivision E3, Bangalore Water Supply and Sewerage Board (BWSSB). Boundary data provided by NextDrop.

The NextDrop system thus generated two types of crowd-sourced data: reported valve opening and closing times from valvemen working *within* the water utility, and accuracy checks from consumers *who do not work for* the utility. While the valvemen data is atypical in the sense that it is sourced from contractors for the utility, the ensuing discussion will show that the standard inferential challenges cited for crowd-sourced data apply more generally.

Given the aforementioned high rates of cell phone penetration in urban India, millions of households could potentially benefit from such a system. NextDrop rolled out this system in the Indian cities of Hubli-Dharwad (population 1 million), Mysore (population 900,000), and Bangalore (population 8 million) with the permission of each city's state-owned utility.[6]

[6] NextDrop was started by a group of U.C. Berkeley engineering graduates, among others.

SURMOUNTING INFERENTIAL AND PRACTICAL CHALLENGES

While crowd-sourced data like NextDrop's offers unprecedented opportunities to analyze social and political phenomena, it presents significant challenges to researchers. First, crowd-sourced data may not describe underlying processes well because data is contributed selectively and reports may be inaccurate. Second, it can be difficult to analyze the impact of crowd-sourced data upon social and political processes—such as protest or political polarization—because crowd-sourced data is often easily shared between individuals. This makes it difficult to distinguish between people who have and have not been exposed to crowd-sourced data, making it difficult to establish a clear counterfactual or control group. Third, the generation and analysis of crowd-sourced data poses significant technical and logistical challenges, such as ensuring data continuity.

In this section, we illustrate these challenges, and present strategies for addressing them. First, concerns regarding descriptive inference, we argue, can be addressed through what geographers, computer scientists, and others utilizing remote sensing technology call "groundtruthing": assessing the accuracy of remotely-collected data, like satellite images, through comparisons with systematic samples of data from the "ground" (Story & Congalton, 1986).[7] Crowd-sourced data can be compared with survey data or data collected systematically through qualitative research, to validate the data generating process, better understand selection biases, and uncover inaccuracies in crowd-sourced

[7] The emphasis on assessing the accuracy of a principle source of remotely-collected data thus differs subtly from triangulation, usually defined as inference based on multiple sources of evidence, such that "diverse viewpoints cast light upon a topic" (Olsen, 2004). Groundtruthing, in contrast, focuses on data validation rather than inference.

data. Further, scholars who wish to examine the causal effects of introducing crowd-sourced data on phenomena like political participation can design experiments that deliberately minimize information sharing or explicitly measure the extent to which it occurs. Finally, interdisciplinary collaborations and partnerships with outside organizations can help tackle important technical and logistical hurdles associated with collecting and analyzing crowd-sourced data. These challenges and strategies are summarized in Table 2 (below). We draw on our own research involving NextDrop's crowd-sourced data on water delivery in urban India—which provided information about the allocation of a scarce resource that even state authorities did not possess—to illustrate these challenges and potential solutions.[8]

Table 2. Strategies for Addressing Difficulties Posed by Crowd-Sourced Data

Problems	Solutions
Description inference • Selection bias • Inaccurate reporting	• Study the selection process explicitly • Qualitative and quantitative groundtruthing • Systematic recruitment of participants
Causal inference • Information-sharing	• Cluster-level rather than individual or household-level randomization • Measuring spillovers
Logistical, technical, and political difficulties	• Partnerships with organizations collecting data • Interdisciplinary collaborations

Strategies Adopted to Address Inferential Challenges

In this section, we describe strategies to help compensate for the inferential difficulties likely to arise when working with crowd-sourced data, using illustrations from our research. Descriptive inference, we show, can be improved through explicitly examining the selection process through which data is contributed and conducting groundtruthing

[8] This section draws upon AUTHOR (Forthcoming) and AUTHOR (2017).

exercises with data collected "on the ground," such as surveys and ethnographic observation. Causal inference regarding the impact of crowd-sourced notifications on citizens could be improved through the use of cluster-level rather than household-level random assignment. Finally, we review strategies for surmounting the major logistical and technical challenges associated with collecting and analyzing crowd-sourced data, such as our water intermittency data, including incentivizing contributions and technical partnerships.

Improving Descriptive Inference: Selection Bias and Correctives

NextDrop's data, like other types of crowd-sourced data, potentially suffered from selection bias. Unlike traditional household surveys, which are based on a well-defined sampling frame, crowd-sourced data is typically collected from a distributed network of users about whom one often possesses limited information. Datasets comprised of crowd-sourced data often suffer from selection bias as a result, especially when data is collected through the internet or smartphones (which can leave out large portions of the population) or when analysts limit their searches to "geo-tagged" data, which includes location information.[9] In the case of the water valve opening and closing reports NextDrop solicited from water valvemen, one main danger was selective reporting: though NextDrop solicited notifications from the entire set of valvemen in areas where it operated, valvemen might not submit notifications for each valve operation. If certain

[9] Only a small percentage of social media data is geo-referenced, because this requires obtaining user consent or extracting location information from posted messages using automated text analysis. For example, approximately 25% of Tweets are geo-tagged (Bryant, 2010; DuVander, Adam, 2010). On the general point of selection bias in crowd-sourced data, see Mayer-Schönberger and Cukier (2013) and Offenhuber (2017, p. 169).

valvemen or groups of valvemen systematically underreported valve openings and closings, one might falsely infer that certain valve areas received less water than they in fact did. Meanwhile, what use would NextDrop water notifications be to consumers if they came on some, but not all, delivery days?

Groundtruthing— i.e., comparing crowd-sourced data with a small sample of systematically collected data from the "ground" in order to gauge accuracy— can illuminate selective patterns of reporting, suggesting where crowd-sourced data may yield biased measures and thus where correctives or adjustments need to be made. Scholars using remote sensing data such as satellite and radio imagery have long assessed data accuracy through groundtruthing. Archeologists, for example, compare satellite imagery of archeological sites with ground surveys of particular areas chosen to verify the interpretation of particular shapes and features (e.g., Hargrave, 2006). Groundtruthing can also be profitably employed to understand the representativeness of crowd-sourced data. Censuses or other surveys following systematic sampling procedures can provide particularly good benchmarks because they do not select respondents based on their ability to report data through crowd-sourcing mechanisms.[10] Researchers can also systematically assess reporting patterns through qualitative research, using interviews and participant observation to gain a better understanding of potential biases in reporting patterns.

We utilized both *quantitative and qualitative groundtruthing strategies* to detect bias stemming from selective reporting in NextDrop's valveman report data. One

[10] Van der Windt and Humphreys (2014), for example, compare conflict data sourced electronically from observers with survey data.

quantitative exercise compared the number of reports submitted by each valveman against the utility's official supply schedule, calculating the number of reports received from each valveman as a percentage of the number of expected reports. While the actual supply schedule deviates from the official supply schedule, such deviations usually occur during a given day; valvemen rarely skip a neighborhood's allotment for a given five-day window. This helped identify valvemen who were not contributing as often as they should be, as well as local utility offices ("service stations") with concentrations of underperforming valvemen. Our analysis of this data suggested that there was significant variation both at the individual level (within a given service station) and across service stations.[11] Note that an alternative approach—less useful for our case, because NextDrop had enrolled the full set of water valvemen in particular areas in their system—is to counter selection bias by deliberately selecting a systematic sample of participants, each of whom possesses the ability and inclination to contribute.[12] Rates of selective reporting would likely be far lower within such a sample.

We also employed qualitative groundtruthing to understand patterns of non-reporting. We selected valvemen across the spectrum of non-reporting rates to interview and "shadow" during their normal work routine.[13] During these shadowing exercises, a

[11] See AUTHOR (Forthcoming) for more detail.
[12] Lawrence (2017), for example, constructed a systematic sample of "first mover" protesters and potential protesters in Morocco. This in turn allowed her to recruit participants for a Facebook survey experiment from a network of activists. Van der Windt and Humphreys (2014) provided a set of individuals in randomly selected villages in the Democratic Republic of Congo with mobile phones and training in reporting conflict events.
[13] Our research focused on nine valvemen in one of the utility's 32 subdivisions. They were shadowed for approximately four months in total.

team member noted whether or not valvemen submitted a report each time they opened and closed a water valve.[14]

Both types of groundtruthing exercises identified individual valvemen who were less likely to contribute reports. We, as well as NextDrop, conducted follow-up research to understand why some valvemen contributed reports less regularly; in other words, we made *the selection process itself an object of investigation*. We used the reporting rates as a point of departure for systematic investigation of the circumstances under which valvemen tend to comply with NextDrop requests for information, triangulating between the large dataset of valvemen reports and ethnographic study of selected valvemen.[15] We aimed to gain a better sense of the individual and community level factors associated with higher rates of reporting. NextDrop, on their side, investigated problems with specific valvemen and service stations, and developed responses to the performance barriers they uncovered in particular cases. These responses included purchasing new sandals for a particularly productive valveman, repairing a valveman's motorbike, taking valvemen out to lunch, etc.[16]

Improving Descriptive Inference: Inaccurate Reporting and Correctives

Related to, but distinct from, selection bias is inaccurate reporting. Among the subset of the population that participates in crowd-sourcing processes, some may report

[14] Observation of this sort can, of course, suffer from the Hawthorne effect. In our case, the danger would be that valvemen would be more likely to report as expected when observed. However, this made observations of divergence from expectations in our presence particularly informative.
[15] AUTHOR (Forthcoming) provides this analysis.
[16] A fuller discussion of the use of incentivizing data contributions appears below.

misleading or inaccurate data. For example, those with grudges might incorrectly report that certain low-ranking officials have engaged in corrupt practices. As with selection, quantitative and qualitative groundtruthing exercises can potentially allow for validation of the accuracy of crowd-sourced data. Surveys based on systematic samples are again particularly useful. One could imagine, for example, comparing the results of a survey utilizing list experiments to measure levels of corruption in a particular location, and then comparing the results with crowd-sourced corruption reports. An alternative, and complementary approach, is qualitative groundtruthing. Participant observation and interviews can bring to light inaccurate reporting, and uncover factors that may contribute to it.

We engaged in both quantitative and qualitative groundtruthing exercises to assess the accuracy of NextDrop's valveman reports. First, we compared the water opening and closing time reports with geo-referenced surveys of households regarding water supply timing. As a part of an impact evaluation of NextDrop's services our research team fielded a two-round 3,000 household survey in Bangalore utility subdivision E3.[17] This survey contained detailed questions regarding the frequency and variability of water services. GPS coordinates were collected for each household, which allowed us to place each in a particular valve area—for which valvemen were responsible for reporting water valve opening and closing times. This allowed us to assess the accuracy of valveman reports. Our qualitative research with the water valvemen (described above) also allowed us to conduct qualitative groundtruthing. A team member noted the timing of water valve openings and closures, and whether or not valvemen sent

[17] Details in this paragraph are drawn from AUTHOR (2017).

notifications immediately to NextDrop, or whether they instead sent them at other times of the day.[18]

Causal Inference: Pitfalls and Solutions

In addition to using crowd-source data to describe phenomena of interest, researchers may want to assess the social or political impacts of introducing crowd-sourcing processes themselves. One can potentially measure the impact of introducing crowd-sourced data by comparing its impact with a similar social context where the application has not been introduced. Finding a good comparison group for the "treatment" group, however, is often challenging because crowd-sourced data is easily disseminated. In the context of an experiment or quasi-experiment, individuals in the control or comparison group may gain access to the treatment. For example, easy access to social media across the area of research makes it challenging to isolate the causal relationship linking use of social media to protest strength. Comparisons between groups with and without access to a crowd-based technology may require a fair amount of geographic separation, which in turn may reduce the comparability of the treatment and control groups.

We grappled with the difficulty of establishing valid counterfactuals as we designed our impact evaluation of NextDrop's services on household welfare, political

[18] Van der Windt and Humphreys (2014) also utilize qualitative groundtruthing, in their case to assess the accuracy of reports of conflict. The authors sent field coordinators to verify the quality of their "crowdseeded" conflict data from the Democratic Republic of Congo: coordinators observed whether or not contributors understood coding schemes and assessed the accuracy of reporting. The paper, unfortunately, does not provide detail on the types of qualitative research methods used to assess data accuracy.

attitudes, and participation. The main inferential challenge we faced was that households receiving information about expected water timings could share the information with neighbors. We verified our initial hunches about the problem of information sharing by conducting a small telephone survey with a systematic sample of existing NextDrop customers in Bangalore, finding that those who thought the service useful said that they shared information about water arrival times with neighbors. However, notification information would only be useful within the same valve area, or the area serviced through the same water valve.

In cases like this, where crowd-sourced data is only useful within a circumscribed geographic area, researchers may be able to prevent information-sharing from affecting study results by employing cluster-level rather than individual-level randomization. Such research designs typically require a large number of clusters to obtain sufficient statistical power, which may or may not be feasible depending on the research setting.

In our study of the household-level impacts of NextDrop's water notifications, we followed this approach, randomizing at the cluster- rather than household-level. Our 120 clusters were separated from one another by a street or two. Because valve areas only encompassed 50-200 households and water notification information was only useful within a valve area, this was sufficient to prevent information-sharing almost entirely. An alternative approach is to measure information spillovers explicitly. In such cases, researchers can randomly assign clusters to different levels of treatment, with varying proportions of individuals in different clusters being eligible for treatment. Researchers then assess the extent of spillover within treatment groups by measuring if untreated members in treated clusters in fact receive treatment, as well as whether or not they exhibit the hypothesized effects of treatment.[19]

Addressing Organizational and Technical Challenges

There are also numerous organizational and technical challenges associated with collecting and analyzing crowd-sourced data, particularly if one wants to amass data consistently over time rather than for a short period. First, the entity or organization interested in collecting data must somehow incentivize contributors effectively over time, especially if contributors do not benefit directly from data collection (Parikh, 2015). Second, organizations typically collect crowd-sourced data using complicated new technologies that require significant technical and organizational infrastructures to develop and administer. In addition, securing approval for the deployment of crowd-sourcing approaches may require delicate negotiations with government officials.

Incentivizing Contributions over Time

Contributions to crowd-sourcing systems must be incentivized since users can often "free ride," governments and other entities may not respond to crowd-sourced feedback, and contributors may not even benefit from the contributed data (World Bank, 2016, p. 164). Incentives can be financial or non-financial, depending on the context. Some organizations pay contributors. Amazon Mechanical Turk, for example, is a crowd-based online marketplace where tasks are completed by "turkers" for an advertised price.

[19] Many control group observations are usually needed for sufficient statistical power under such a design (Baird, Bohren, McIntosh, & Ozler, 2015; Gerber & Green, 2012, p. 260).

Similarly, researchers have offered internet café and phone credits to encourage participation in studies involving internet usage and text messaging (Bailard, 2012; Grossman, Humphreys, & Sacramone-Lutz, 2014). Meanwhile, organizations such as Wikipedia, Waze, and Twitter rely on non-financial incentives that focus on building community norms or appealing to contributors' self-images to encourage data reporting (Parikh, 2015).

The NextDrop case also illustrates the importance of establishing robust organizational and technical infrastructure for incentivizing contributions of crowd-sourced data. Recognizing that its system benefitted households who received notifications as well as the water utility's management more than the valvemen, NextDrop experimented with several means of incentivizing valvemen to contribute accurate data on a regular basis. In its first location, the medium-sized Indian city of Hubli-Dharwad, NextDrop experimented with a variety of incentive schemes involving in-kind rewards for valvemen. NextDrop claimed that this first approach secured reasonably high rates of valveman participation (AUTHOR, 2018). NextDrop judged this method to be too labor-intensive, and therefore adopted an alternative approach as they scaled-up in Bangalore. There, NextDrop entered a formal MOU with the state water utility, which instructed the valvemen's supervisors to cooperate with NextDrop. NextDrop delivered reports to valvemen's supervisors on a weekly basis that detailed individual reporting rates, giving supervisors the information needed to enforce compliance. Our groundtruthing exercises in Bangalore revealed that this new approach did not work effectively: in our study area, approximately 70% of valvemen submitted notifications regularly, but at least two-thirds of the notifications submitted were inaccurate. These findings illustrate the importance of instituting effective schemes that not only elicit information, but *correct* information. NextDrop's efforts to obtain citizen feedback on the water notifications proved even less effective, as they failed to solicit feedback regularly and when they did so, they received very few responses.

Surmounting Technical, Organizational, and Political Challenges

Collecting and analyzing crowd-sourced data pose major technical, organizational, and political challenges for social scientists. Crowd-sourcing applications handle large datasets, which are difficult to collect and analyze using traditional techniques. In some cases, it may be necessary to obtain political or regulatory approval to work with crowd-sourced data. In addition, the exact collection and storage protocol followed by third parties, such as Twitter or Facebook, may not be public information—even if some crowd-sourced data is publicly accessible. Social scientists have employed two strategies to address these technical and logistical challenges: interdisciplinary collaborations with colleagues in computer science, civil engineering, and information schools;[20] and research collaborations with organizations collecting crowd-sourced data, such as Facebook and Google (e.g., Bond & Messing, 2015).

In the context of our research, partnering with NextDrop allowed us to overcome numerous technical and political hurdles. NextDrop's organizational and technical capacity enabled it to surmount the significant technical challenges associated with

[20] Centers facilitating such collaborations include the Social Media and Political Participation Laboratory at New York University (http://smapp.nyu.edu/about.html), the Center for Information Technology Research in the Interest of Society at University of California at Berkeley (http://citris-uc.org/about-citris/), and the Media Cloud at Harvard and MIT (http://mediacloud.org/).

collecting and disseminating the valvemen data. The firm's software developers and field teams developed and executed systems for physically mapping hundreds of valve areas using GPS, collecting GPS coordinates from thousands of households, and collecting notifications from valvemen and feedback from households in an automated fashion. This was a major undertaking given that existing water system maps were out-of-date and inaccurate, especially in informal settlements. GPS coordinates were difficult to collect due to tall buildings, narrow streets, and weak data network connectivity; and the multiple languages spoken by city residents did not always display well on mobile devices. Needless to say, our small team of university-based social scientists was not equipped to develop and maintain such technical and organizational infrastructure. Partnering with NextDrop in this research also provided us with access to the crowd-sourced valveman reports, which allowed us to examine biases and inaccuracies in the data. NextDrop also secured the political approval for our impact evaluation, which involved convincing authorities to delay their rollout in a section of the city and only offer services to our designated treatment group.

WHAT CROWD-SOURCED DATA TELLS US ABOUT PRINCIPAL-AGENT PROBLEMS IN URBAN SERVICE DELIVERY

What were we ultimately able to learn about the politics of urban service delivery in the Global South—especially service delivery in informal settlements—through our analysis of NextDrop's crowd-sourced data? The water timing data NextDrop collected from Bangalore's valvemen ultimately proved most useful for understanding relationships between water valvemen and the utility. Our Bangalore survey data suggests that the individuals charged with managing household water supply—typically women—

rarely turned to neighborhood leaders or politicians regarding water complaints. Over 60% of our respondents in low-income neighborhoods reported contacting water valvemen with service problems, while less than 10% contacted a local leader or city councillor. This suggests that political scientists should devote more energy to examining the everyday interactions between frontline workers and citizens, and the extent to which frontline workers like valvemen exercise significant discretion in such interactions.[21] The activities of street-level bureaucrats (SLBs) has received far too little attention in recent comparative politics research on slum politics, which focuses on other types of intermediaries such as local leaders and clientelistic brokers (e.g., Auerbach, 2016; Jha, Rao, & Woolcock, 2007).

NextDrop's crowd-sourced valvemen reports on water opening and closing times provided a means of examining principal-agent problems within the water utility's complex bureaucracy. Inaccuracies and missingness found when comparing NextDrop's valveman report data with the official water supply schedule and our surveys revealed significant noncompliance with central mandates. While Bangalore's water utility management redefined their valvemen's job descriptions so as to include sending valve notifications to NextDrop, a large number of valvemen reported inaccurate information or did not report at all. Water valvemen clearly exercise significant discretion, both in terms of how they respond to mandates from utility management and in terms of how they respond to local communities. Our parallel qualitative fieldwork revealed that valvemen perceived sending notifications as an additional burden on top of existing "core"

[21] While recent work suggests that citizens most often approach state officials, such as elected representatives, directly (e.g., Kruks-Wisner, 2011; Bussell, 2017; Kruks-Wisner, 2018), our emphasis here on SLBs is distinct.

responsibilities, such as turning water valves on and off and assisting with network repairs (AUTHOR, 2018). These findings suggest that the political science and economics literature on public goods provision should pay far more attention to SLB discretion and how SLB actions ultimately affect service allocations, quality, and citizen perceptions of the state. Crowd-sourced data like NextDrop's can help shed light on these relationships.

A second finding is that valvemen varied in their compliance with the utilities' directive to work with the NextDrop system (AUTHOR, 2018). In other words, principal-agent problems appeared to affect some valvemen more than others. Triangulating between the valveman reports, an original dataset on valveman and valve area characteristics, and qualitative research, we observed that valveman compliance was strongly associated with non-political factors, such as individual and family financial circumstances. Valvemen struggling to meet family obligations and needing to earn additional money through odd jobs complied at lower rates. While the public administration literature has studied how individual characteristics, such as education and social identity, affect the performance of SLBs, it has failed to consider such personal financial and familial circumstances (AUTHOR, 2018). More broadly, our analysis suggests that the political science literature on intermediaries such as brokers may pay insufficient attention to how individual characteristics of this sort affect intermediary behavior.

While our analyses provided important insight into the extent and drivers of noncompliance with organizational directives, it also raises questions for future research. To what extent, for example, do SLBs like valvemen exercise autonomy in their decisions regarding to whom to allocate water or other services, and when? To what extent do they succumb to pressure from neighborhood leaders or local politicians, becoming part of clientelistic machines? Ethnographic accounts of Mumbai's water sector suggest that city councillors do pressure utility employees to direct water towards particular constituencies (e.g., Anand, 2012; Björkman, 2015, p. 161), yet they also emphasize the extent to which water valvemen derive autonomy from significant information asymmetries (see Björkman, 2015). While our survey data suggests that citizens rarely if ever approach anyone other than valvemen with water problems, some valvemen did report that local politicians contacted them. Understanding how SLBs balance the expectations and demands of multiple "principals"—including supervisors, local politicians, and citizens—is an important avenue for future research. Such research could draw on multiple types of information, including crowd-sourced data.

CONCLUSION AND IMPLICATIONS

Urban service delivery in the Global South is extremely important but very difficult to study. The bureaucracies charged with providing basic necessities, such as transportation, infrastructure, water and sanitation, and policing, possess complex hierarchies and underlying service delivery infrastructure. We currently know little about the internal workings of these bureaucracies or how citizens experience them. Yet such knowledge is fundamental to understanding why some states cope more effectively than others with urban growth—extending services to new populations—while others do not.

Until recently, it has been extremely difficult to obtain quality data on urban service delivery in the Global South. Crowd-sourced data constitutes a new source of information about patterns of allocation and service quality, as well as about the activities of employees of the bureaucracies delivering these services. Reasonably accurate data submitted by utility employees or citizens can provide information about patterns of allocation and service quality when service providers do not themselves possess accurate information. It can also provide governments with a means of obtaining citizen feedback on spending priorities. Even inaccurate data supplied by employees at the behest of management provides a means of examining principal-agent problems in urban service providers.

Crowd-sourced data also promises to provide analytic leverage on a host of other questions of interest to political scientists. Scholars have already begun to exploit this data to examine phenomena like protest and online forms of political participation, political conflict, and public opinion. Relatedly, scholars have begun to use crowd-sourced data to better understand undocumented populations, such as unregistered refugees. There are also numerous areas in which businesses and other entities are using such data that our discipline has yet to exploit. For example, data from crowd-sourcing platforms focused on corruption, traffic conditions (including police presence), and crime have received very little scholarly attention.

This paper has outlined strategies for addressing some of the main pitfalls to be avoided when pursing these exciting analytic opportunities. First, scholars must consider various threats to descriptive inference. Data may be contributed selectively, and contributed data may be inaccurate. Each of these problems can lead to biased descriptions of the phenomena under investigation. Scholars can engage in groundtruthing exercises to identify selection processes and assess measurement error by utilizing surveys involving systematic sampling and qualitative research. We expect these strategies to be useful across a wide range of subject and geographic areas.

Scholars interested in examining the impact of crowd-sourcing processes, such as social media usage on social and political phenomena, face another set of challenges. The fact that technology users can share information very easily means that it can be difficult to locate a reasonable comparison group. Drawing on studies of information-sharing in other contexts, we discuss two main approaches to addressing this problem: cluster-randomized experiments when information-sharing is limited to specific geographic areas, and experimental approaches that explicitly measure the extent to which information-sharing occurs.

Most formidable, perhaps, are the organizational and technical challenges associated with working with crowd-sourced data. Collecting and analyzing such data requires very high levels of technical sophistication, and often a well-developed organizational apparatus that provides appropriate incentives for contributors. It may also be necessary to develop and nurture political relationships to ensure data access. These challenges can be addressed, at least in part, through interdisciplinary collaborations or by partnering with organizations that dedicate themselves to collecting such data.

Works Cited

Anand, N. (2012). Municipal disconnect: On abject water and its urban infrastructures. *Ethnography*, *13*(4), 487–509.

Auerbach, A. (2016). Clients and Communities: The Political Economy of Party Network Organization and Development in India's Urban Slums. *World Politics*, *68*(1), 111–148.

AUTHOR. (Forthcoming). Frontline Worker Compliance with Transparency Reforms: Barriers Posed by Family and Financial Responsibilities. *Governance*.

AUTHOR. (2017, Fall). *Flows, Leaks, and Blockages in Informational Interventions: A Field Experimental Study of Bangalore's Water Sector*. Berkeley, CA.

Bailard, C. S. (2012). A Field Experiment on the Internet's Effect in an African Election: Savvier Citizens, Disaffected Voters, or Both? *Journal of Communication*, *62*(2).

Baird, S., Bohren, J. A., McIntosh, C., & Ozler, B. (2015). Designing Experiments to Measure Spillover Effects, Second Version. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2619724

Barberá, P. (2013, February). *The 2013 Italian Parliamentary Elections on Twitter*.

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, *23*(1), 76–91.

Barberá, P., & Metzger, M. (2013). *A Breakout Role for Twitter? The Role of Social Media in the Turkish Protests* (Social media and political participation lab data report). Social Media and Political Participation Lab.

Björkman, L. (2015). *Pipe politics, contested waters: Embedded infrastructures of millennial Mumbai*. Durham, NC: Duke University Press.

Bond, R., & Messing, S. (2015). Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook. *American Political Science Review*, *109*(01), 62–78.

Boutet, Antoine, Hyoungshick Kim, and Eiko Yoneki. (2012). What's in Your Tweets? I Know Who You Supported in the UK 2010 General Election. *ICWSM*.

Breuer, A., Landman, T., & Farquhar, D. (2012). *Social Media and Protest Mobilization: Evidence from the Tunisian Revolution* (SSRN Scholarly Paper No. ID 2133897). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2133897

Bryant, M. (2010). Twitter Geo-fail? Only 0.23% of tweets geotagged. Retrieved April 20, 2015, from http://thenextweb.com/2010/01/15/twitter-geofail-023-tweets-geotagged/

Bussell, J. (2017, April). Serving Clients and Constituents: Experimental Evidence on Political Responsiveness.

Calvo, E. (2015). *Anatomia Politica de Twitter en Argentina*. Buenos Aires, Argentina: Capital Intelectual.

Carlson, M., Jakli, L., & Linos, K. (Forthcoming). Rumors and Refugees: How Government-Created Information Vacuums Undermine Effective Crisis Management. *International Studies Quarterly*.

Ching, A., Zegras, C., Kennedy, S., & Mamun, M. (2013). A User-Flocksourced Bus Experiment in Dhaka: New Data Collection Technique with Smartphones.

Transportation Research Record: Journal of the Transportation Research Board. Retrieved from http://web.mit.edu/czegras/www/Flocksource_JUT.pdf

DuVander, Adam. (2010). 3 Reasons Geocoded Tweets Haven't Caught On and 2 Reasons Not to Worry. Retrieved April 20, 2015, from http://www.programmableweb.com/news/3-reasons-geocoded-tweets-havent-caught-and-2-reasons-not-to-worry/2010/02/17

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200.

Furtado, V., Caminha, C., Ayres, L., & Santos, H. (2012). Open Government and Citizen Participation in Law Enforcement via Crowd Mapping. *IEEE Intelligent Systems*, *27*(4), 63–69.

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton.

Grossman, G., Humphreys, M., & Sacramone-Lutz, G. (2014). "I wld like u WMP to extend electricity 2 our village": On Information Technology and Interest Articulation. *American Political Science Review*, *108*(03), 688–705.

Hargrave, M. L. (2006). Ground Truthing the Results of Geophysical Surveys. In J. K. Johnson, M. Giardano, K. L. Kvamme, R. B. Clay, T. J. Green, & R. A. Dalan (Eds.), *Remote Sensing in Archaeology* (pp. 269–303). Tuscaloosa, AL: University of Alabama Press.

Iliffe, M., Sollazzo, G., Morley, J., & Houghton, R. (2014). Taarifa: Improving Public Service Provision in the Devel- oping World Through a Crowd-sourced Location Based Reporting Application. *OSGeo Journal*, *13*(1), 34–40.

Jamal, A., Keohane, R., Romney, D., & Tingley, D. (2015). Anti-Americanism or Anti-Interventionism in Arabic Twitter Discourses. *Perspectives on Politics*, *13*(1), 55–73.

Jha, S., Rao, V., & Woolcock, M. (2007). Governance in the Gullies. *World Development*, *35*(2), 230–246.

Klopp, J., Mutua, J., Orwa, D., Waiganjo, P., White, A., & Williams, S. (2014). Towards a Standard for Paratransit Data: Lessons from Developing GTFS Data for Nairobi's Matatu System. Presented at the Transportation Research Board 93rd Annual Meeting. Retrieved from http://trid.trb.org/view.aspx?id=1289853

Kruks-Wisner, G. (2011). Seeking the Local State: Gender, Caste, and the Pursuit of Public Services in Post-Tsunami India. *World Development*, *39*(7), 1143–1154.

Kruks-Wisner, G. (2018). The Pursuit of Social Welfare: Citizen Claim-Making in Rural India. *World Politics*, *70*(1), 122–163. https://doi.org/10.1017/S0043887117000193

Lawrence, A. K. (2017). Repression and Activism among the Arab Spring's First Movers: Evidence from Morocco's February 20th Movement. *British Journal of Political Science*, *47*(3), 699–718.

Lipsky, M. (1980). *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. New York, NY: Russell Sage Foundation.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.

Offenhuber, D. (2017). *Waste is Information: Infrastructure Legibility and Governance*. Cambridge, MA: The MIT Press.

Olsen, W. (2004). Triangulation in Social Research: Qualitative and Quantitative Methods Can Really be Mixed. In M. Holborn (Ed.), *Developments in Sociology*. Causeway Press.

Parikh, T. (2015). Digital Data Collection for improving service delivery: A framework for decision-makers.

Peixoto, T., & Fox, J. (2015). When Does ICT-Enabled Citizen Voice Lead to Government Responsiveness. *2016 World Development Report Background Paper*, (Internet for Development).

Post, A. E., Bronsoler, V., & Salman, L. (2017). Hybrid Regimes for Local Public Goods Provision: A Framework for Analysis. *Perspectives on Politics*, *15*(4), 952–966.

Poushter, J. (2016, February 22). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Retrieved June 12, 2017, from http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/

Raza, D. (2017, June 10). 'I saw it on WhatsApp': Why people believe hoaxes on the messaging app. Retrieved June 12, 2017, from http://www.hindustantimes.com/i-saw-it-on-whatsapp-why-people-believe-hoaxes-on-the-messaging-app/story-TTAJjgLC7eL2Mb0LNxVGuJ.html

Sachdev, C. (2017, June 7). If You See Dirty Water, Don't Just Gripe. Talk to the Cloud! [National Public Radio]. Retrieved June 9, 2017, from http://www.npr.org/sections/goatsandsoda/2017/06/07/527898124/if-you-see-dirty-water-dont-just-gripe-talk-to-the-cloud

Scott, J. C. (1996). State Simplifications: Nature, Space, and People. *Nomos*, *38*, 42–85.

Starbird, K., & Palen, L. (2012). (How) Will the Revolution Be Retweeted?: Information Diffusion and the 2011 Egyptian Uprising. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 7–16). New York, NY, USA: ACM.

Story, M., & Congalton, R. G. (1986). Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing*, *52*(3), 397–399.

Telecom Regulatory Authority of India. (2017). *The Indian Telecom Services Performance Indicator Report*. New Delhi, India. Retrieved from http://www.trai.gov.in/

Touchton, M., & Wampler, B. (2014). Improving Social Well-Being Through New Democratic Institutions. *Comparative Political Studies*, *47*(10), 1442–1469. https://doi.org/10.1177/0010414013512601

UNICEF. (2012a). U-report application revolutionizes social mobilization, empowering Ugandan youth. Retrieved April 16, 2015, from http://www.unicef.org/infobycountry/uganda_62001.html

UNICEF. (2012b, August 16). TIME Magazine Covers UNICEF supported mTrac system in Uganda. Retrieved April 16, 2015, from http://unicefstories.org/2012/08/16/time-magazine-covers-unicef-supported-mtrac-system-in-uganda/

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. A. (2015). Political Expression and Action on Social Media: Exploring the Relationship Between Lower- and Higher-Threshold Political Activities Among Twitter Users in Italy. *Journal of Computer-Mediated Communication*, *20*(2), 221–239.

van den Berg, C., & Danilenko, A. (2011). *The IBNET Water Supply and Sanitation Performance Blue Book*. Washington D.C.: The World Bank.

van der Windt, P., & Humphreys, M. (2014, February 13). *Crowdseeding Conflict Data*.

World Bank. (2016). *Digital Dividends: World Development Report 2016*. Washington D.C.: The World Bank Group.

Yadav, T. author has posted comments on this articleAnkit. (2015). Tension grips Rithora over objectionable WhatsApp post. Retrieved October 16, 2015, from http://timesofindia.indiatimes.com/city/bareilly/Tension-grips-Rithora-over-objectionable-WhatsApp-post/articleshow/47389539.cms