Available online at www.sciencedirect.com

ScienceDirect

Journal of Urban Management 4 (2015) 74–91

www.elsevier.com/locate/jum

Research Article

Open data for informal settlements: Toward a user's guide for urban managers and planners

Arnab Chakraborty[a,*], Bev Wilson[a], Saket Sarraf[b], Arnab Jana[c]

[a]*University of Illinois at Urbana-Champaign, Champaign, IL, USA*

[b]*ps Collective, Ahmedabad, Gujarat, India*

[c]*Indian Institute of Technology Bombay, Mumbai, Maharashtra, India*

Received 31 October 2015; received in revised form 2 December 2015; accepted 2 December 2015

Abstract

Informal settlements exist in a legally contested space and the quality of – and access to – information about them has been historically limited. The open data movement promises to address this gap by offering alternative sources for information and free or low cost analytical platforms. However, in order to use open data effectively, urban managers and planners need guidance to navigate these new data sources, software, and server platforms, as well as acquire the necessary skills. In this paper, we begin to address these issues by developing a framework that organizes the sprawling and rapidly evolving world of open urban data. Our framework includes three broad categories (1) inputs and resources, (2) activities and outputs, and (3) outcomes. We then identify and describe the key subcomponents under each, and list the prominent products and resources available to urban managers and planners. For example, under inputs and resources, we discuss open urban data sources such as Open Street Maps, cyberinfrastructure for web hosting, application deployment, and data processing, and open source software that can be used to analyze and visualize collected or derived data. We also identify the key resources available to planners for training and discuss the complementary opportunities presented by conventional datasets such as census and open urban data. Finally, using examples from ongoing activities in Mumbai, we show how open data resources can be useful for understanding urbanization and better integrating informal settlements into formal urban management and planning processes. We suggest that urban managers and planners working in informal settlements should take greater advantage of open data resources in order to both better address current challenges as well as for shaping a better future for the communities they serve.

Keywords: Open data; Informal settlements; Slums; Planning support; Nongovernmental organizations

1. Introduction

Urban data is a term broadly applied to information characterizing the many facets of human settlements, including socio-economic and demographic makeup of its residents, infrastructure and environmental conditions, and historic trends in

*Correspondence to: 611 E Lorado Taft Drive, 111 Temple Buell Hall, MC619, Champaign, IL 61820, USA.

E-mail addresses: arnab@illinois.edu (A. Chakraborty), bevwilso@illinois.edu (B. Wilson), saket.sarraf@gmail.com (S. Sarraf), arnab.jana@iitb.ac.in (A. Jana)

Peer review under responsibility of Zhejiang University and Chinese Association of Urban Management

http://dx.doi.org/10.1016/j.jum.2015.12.001

growth and change (Batty, 2008). High quality urban data for use as a fact base is an essential component of decision-making at large and small scales, and across a variety of sectors (White & Engelen, 2000, Klosterman, 1994). From an urban management perspective, data can be useful for identifying areas for infrastructure investments (Gramlich, 1994), pinpointing locations of environmental and public health hazards (Lawson et al., 1999), conserving biodiversity (Maddock & Du Plessis, 1999), and delivering basic as well as emergency services (Lee, 2007).

However, not all urban areas are created equal when it comes to the quality and availability of data for urban management. Detailed and standardized datasets are traditionally produced by public agencies, which often control its type, storage, and dissemination. In developing countries in particular, this data is often difficult to access for non-governmental agencies and researchers (Rambaldi, Kyem, McCall, & Weiner, 2006). The quality of – and access to – urban data is especially limited in areas of informal settlements (Zetter & De Souza, 2000), which exist in a legally contested space. As a result, agencies serving these areas often use ad-hoc datasets (Dagdeviren & Robertson, 2009) that can limit the perceived legitimacy of their analyses and findings. On the other hand, formal development plans and other instruments of the urban management agencies pay little attention to the needs of slums, either opting to relegate them to special purpose agencies in charge of "rehabilitation" or using the lack of data to justify their exclusion from formal planning processes and management efforts (Werlin, 1999). Indeed, many have attributed the lack of proactive planning for slums and the continuing challenges faced by public and non-profit sector planners to such attitudes (Werlin, 1999, Das & Takahashi 2009).

Open urban data or free, accessible, and user-generated information (Liu et al., 2015) has recently burst into the scene with the potential to address the gaps in the fact base revealed by relying solely on government-produced and other conventional authoritative datasets (Crooks et al., 2014). Their role in urban planning and management in particular has been touted by both the technology sector and academic researchers (Liu et al., 2015, Crooks et al., 2014). However, in order to realize the promise of open data, urban managers and planners need more detailed guidance on how to navigate a sprawling field of often unrepresentative and unstandardized datasets, varied software and server platforms, and gaps not readily addressed by existing open urban data resources. Navigating these issues is particularly challenging when organizations are operating with fewer resources and lower planning capacity (Donovan, 2012).

In this paper, we begin to address these issues by developing a framework to help urban managers and planners navigate the sprawling and rapidly evolving world of open urban data. Our framework is organized into three broad categories using the conventions of the logic model (Kellogg Foundation, 2004). These categories are: (1) *inputs and resources*, (2) *activities and outputs*, and (3) *outcomes*. We then identify and describe the key subcomponents under each, and list the prominent products and resources available to urban managers and planners. For example, under inputs and resources, we identify *open urban data* sources such as Open Street Maps, *cyberinfrastructure* or platforms for web hosting, application deployment, and data processing, and *open source* software that can be used to analyze and visualize collected or derived data. We also identify the key resources available to planners for training and discuss the complementary opportunities presented by open urban data and conventional datasets such as census. Finally, using examples from ongoing activities in Mumbai, we show how open urban data resources can be useful for understanding urbanization and better integrating informal settlements into formal urban management and planning processes. We suggest that urban managers and planners working in informal settlements should take greater advantage of open data resources in order to both better address current challenges as well as for shaping a better future for the communities they serve.

2. Motivation

We live in an increasingly urbanized world and this trend is expected to continue well into the future, but the locus of urban population growth has shifted from the developed economies of the Global North to developing nations of Asia and Africa (United Nations, 2014a). In many of the fastest growing cities, urban informality, understood as "a state of deregulation, one where the ownership, use, and purpose of land cannot be fixed and mapped according to any prescribed set of regulations or the law" (Roy 2009, p. 80) is an indisputable fact of life, but is often ignored by formal, state-sanctioned planning processes (Patwa, 2013; Pinto, 2015). An example can be seen in the map (see Fig. 1) taken from the Development Plan for Greater Mumbai 2014–2034 prepared by Municipal Corporation of Greater Mumbai, one of two planning agencies with jurisdiction over the city (Chakraborty, Wilson & Bin Kashem, 2015).

Informal settlements, or slums as they are often called, also comprise a significant and increasing proportion of residential dwellings in cities across the developing world. As of 2003, an estimated 1 billion people lived in slums (UN-Habitat 2003). Although the definition of slum varies from country to country and in the case of India, from state to state (Nolan, 2015), the number of people living in slums is expected to increase in the short term for both the nation as a whole and the state of Maharashtra (Fig. 2). In Mumbai, more than one-half of the city's population lived in slums according to the 2001 census, but this figure decreased to 41.9% by 2011 (Bardhan, Sarkar, Jana, & Velaga, 2015).

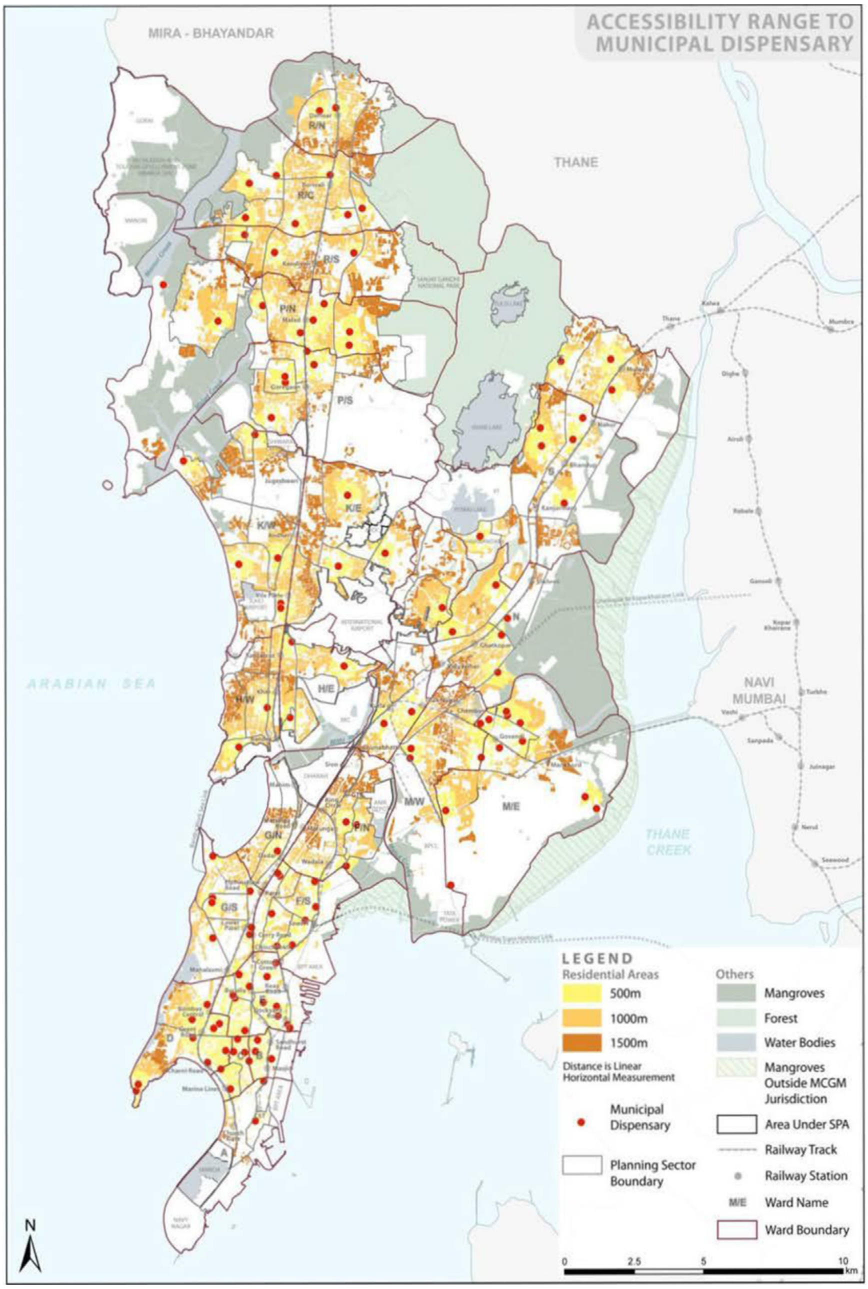

Fig. 1. Map of "Accessibility and Spatial Distribution of Medical Amenities" in Mumbai ignores slums. Red dots are locations of municipal dispensaries or clinics. Shades of Brown show their accessibility to mapped residential areas. Un-shaded areas fall under Special Planning Area, which includes areas under the Slum Rehabilitation Agency. (Source: MCGM, 2014, Existing Land Use Map 17).

The challenges faced by slum residents are extensive and have been well documented (UN-Habitat, 2003) and range from the familiar communicable diseases (Sclar, Garau, & Carolini, 2005) and sanitation (McFarlane, Desai, & Graham, 2014) to displacement (Doshi, 2013). However, neither these challenges nor the inextricable economic and social linkages between slums and formal housing neighborhoods in cities like Mumbai have been able to bridge the disconnect between planning and urban management within these two spheres. While the prevailing view in the developed world has been that informality signals the failure of planning and urban management (UN-Habitat, 2003), it has also been argued that the informality that characterizes Indian cities is an intentional deregulation of urban spaces and processes where "the seeming withdrawal of regulatory power creates a logic of resource allocation, accumulation, and authority" (Roy, 2009, p. 83). By not directly addressing informality, planning efforts give tacit approval and encouragement to their expansion and proliferation in a way that affords considerable flexibility in how these areas are managed. In India "there is no explicit national policy on tenure regularizaton" (Banerjee, 2002, p. 39) which means that slum residents occupy a tenuous position in cities like Mumbai where large-scale evictions allow these areas to function as place holders for future formal-sector development (Arputham & Patel, 2010; Doshi, 2013) as well as reservoirs of flexible and low-cost labor (Bardhan et al., 2015).

Because informal settlements have largely been ignored in formal planning processes, efforts to administer these areas and to provide adequate public services are constrained by a lack of information and data (Clarke, 1991; Abbott, 2001). These limitations have in the past been sidestepped by relying on data acquired through remote sensing (Musakwa & Van Niekerk 2015), but new approaches and technologies are making it increasingly possible to build more robust and comprehensive spatial data infrastructures (SDI) in developing countries (Budhathoki, Bruce, & Nedovic-Budic, 2008). The advent of cyberinfrastructure, open source software, mobile phone applications, and the volunteered geographic information (VGI) has fundamentally reshaped the way scholars and practitioners interact with data and holds considerable promise for improving the factual basis of urban management in cities like Mumbai where informality is an essential characteristic of the urbanization process (Bardhan et al., 2015).

3. Informality and data driven urban management

In Mumbai and many other large cities of the developing world, rather than a centralized government or hierarchy of jurisdictions providing public goods and services in a coordinated fashion, an ad hoc patchwork of often privately constructed infrastructure exists alongside the products of a more distributed and autonomous form of governance

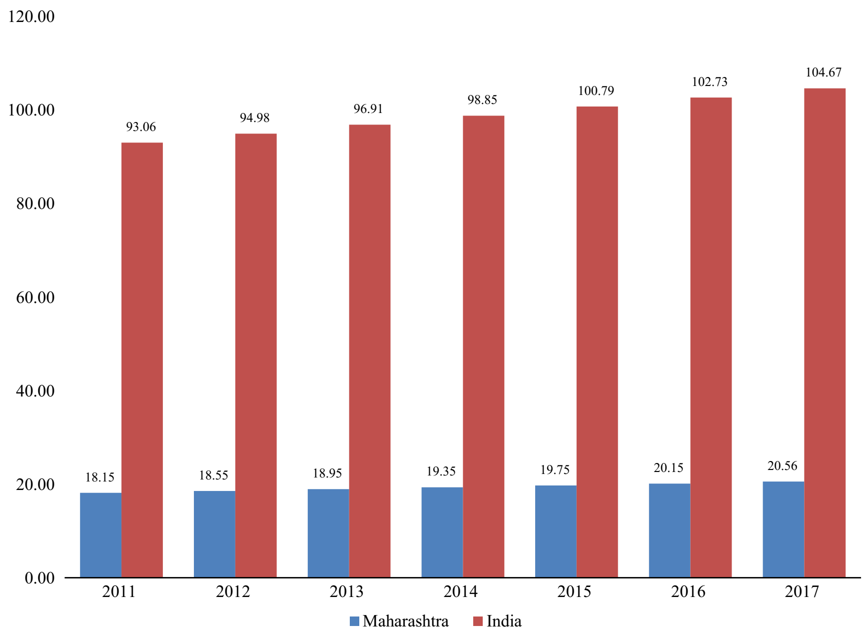

Fig. 2. : Projected slum population for Maharashtra and India, 2011–2017 (Government of India, 2010).

described as “slum entrepreneurialism” (McFarlane, 2012). In many cases, nonprofit organizations or voluntary agencies (VOLAGs) take the lead in supplementing the efforts of slum residents to meet their own needs with respect to infrastructure, public health, and education and function as a liaison between these marginalized populations and the formal sectors of society (Sen, 1999). However, most Indian nonprofit organizations tend to focus on a single issue (e.g., housing, public health) which contributes to the fragmentation of resources and authority that frustrates attempts to better plan for and manage informal settlements (Zérah, 2008; Weinstein, 2014).

The open urban data paradigm (Liu et al., 2015) represents an opportunity to facilitate closer coordination within the community of nonprofit organizations currently working on service delivery in Mumbai's informal settlements, remove a key obstacle to the integration of informal settlements into urban management efforts, and advance research into the role of informality in urbanization processes. New and different types of data and data generating techniques will be a necessary component of this integration. In the contemporary city characterized by “unequal access to information, political power, and other resources, data are a powerful source of influence” (Elwood, 2008, p. 83) and can be understood as a basis for citizen empowerment and wider participation in policy and planning decisions. The work of activist–architect PK Das related to mapping of slums (Das, 2011) and the broader Open Mumbai project (Das, 2012) of which these maps are a part exemplify this perspective. On the other hand, given that “the actors producing knowledge recognized generally in urban planning and management are mainly government working with the private sector” (Baud, Scott, Pfeffer, Sydenstricker-Neto, & Denis, 2014, p. 231) the open urban data paradigm also could be viewed as a threat to existing power structures, which could also help to explain the reluctance of many public agencies and non governmental profit organizations to share the data they collect. The fierce criticism and subsequent withdrawal recently of the draft Mumbai Development Plan (Venkatraman, 2015) underscores this point.

Open urban data could also be conceptualized within the system of plans (SoP) approach established by Hopkins (2001) where not only plans, but information itself could be interpreted as signaling not only the intentions of actors but also as revealing the way that urban processes and issues are framed and understood by those creating and disseminating data. In this way, open urban data could in theory, provide a common fact base for discourse, negotiation, greater understanding, and cooperation among actors.

Finally, most open urban data resources are applicable to both formal and informal settlements but, as we note throughout the paper, informal settlements do present some unique challenges and opportunities, particularly due to the lack of comparable resources and planning capacity in them. Under such condition, open data can offer an alternative to other data sources and should be considered carefully. We should also note that while we tried to focus the examples on informal settlement applications, a lot of the information presented here could also apply to formal settlements.

4. Navigating open data resources: a logic model

Recent advances in open data promise a number of benefits for urban analysis, especially with regard to serving needs of slums. These advances can assist with collecting and analyzing information on urban conditions as well as improving on-the-ground outcomes. Such datasets can also be complementary to traditional data sources and analytical approaches, and offer ways to connect slum management to broader urban management. Despite these advances, there is little guidance for planners on how to navigate this growing array of data sources, and where to get the necessary skills to put the data to appropriate use. This paper attempts to address a part of this gap.

In this section, we organize the available resources into components of a larger framework or *a logic model* with the aim to connect the dots of this sprawling and fast evolving open data world. A logic model is both a tool and set of conventions for visually communicating important relationships between resources, activities, and outcomes (Kellogg Foundation, 2004). Although most frequently associated with program evaluation (Millar, Simeone, & Carnevale, 2001), logic models can also be used in a broader sense as a conceptual framework (Turley et al., 2013) that describes how existing and emerging data, tools, and processes can be linked to strategically support desired outcomes. When presented in this way, the logic model acts as both a practical roadmap and a normative vision for urban management and planning (Funnell & Rogers 2011).

While our framework is geared towards the community of stakeholders currently working on planning and service delivery in informal settlements in Mumbai, the general framework can be more broadly relevant for any kind of urban data analysis. By attempting to integrate these loose strands in urban data resources, our goal is to also remove a key obstacle to the integration of informal settlements into urban management efforts; and advance research into the role of informality in urbanization processes. We foresee such integration to lead to increased availability and sharing of data across sectors; adoption of urban management practices that better integrate informal settlements; and enhance the contributions to the scholarly literatures on urban management and urban informality.

Fig. 3 depicts our conceptual framework for integrating urban informality into urban management that is informed by the principles and conventions of logic models. We adopt the convention of organizing the framework into three areas: (1) *inputs and resources*, (2) *activities and outputs*, and (3) *outcomes*, each with a set of subcomponents. In the rest of this section we briefly introduce these areas, which are then discussed in more detail in the following sections.

The inputs and resources category includes *traditional urban data* sources, such as those maintained by governmental agencies that conduct census surveys and remote sensing, and new *open urban data* sources, that may include portals for volunteered geographic information such as Open Street Maps or Wikimapia. Open sources or resources may also include smartphone apps that collect primary data such as Open Data Kit, and social media APIs such as Twitter, Waze and Facebook, whose datasets can be mined for relevant user generated content information. While some open data systems are not yet fully usable in the slum context (Goel, 2015), their reach is growing, and they offer complementary benefits and address a number of limitations of larger traditional systems. Other components of inputs are *cyberinfrastructure* or platform for web hosting, application deployment, and data processing, and *open source* software that can analyze and visualize collected or derived data. The traditional data sources and such as national census programs and satellite imagery can be costly, updated less frequently, and are not equipped to adequately classify unauthorized land-use or informal settlements. This is particularly challenging in Mumbai where slums grow and change at a fast rate.

Planners also need education in a variety of analytical and visualization techniques that employ the *inputs*. In *activities and outputs* section, we discuss a few techniques we have employed for using open data. We also discuss a number of resources such as forums, online tutorials, and traditional educational modules that the stakeholders can leverage to acquire or hone these skills. Finally, the needs for skills and training will be partly informed by the theories of the field and will likely vary by the role of the stakeholder, say, a research center or a grassroots organization. In the concluding section focused on *outcomes*, we draw some suggestions about how the stakeholders may combine the available inputs and resources in a variety of contexts for both better service delivery within slums as well as better consideration of slums in the larger urban management and planning processes.

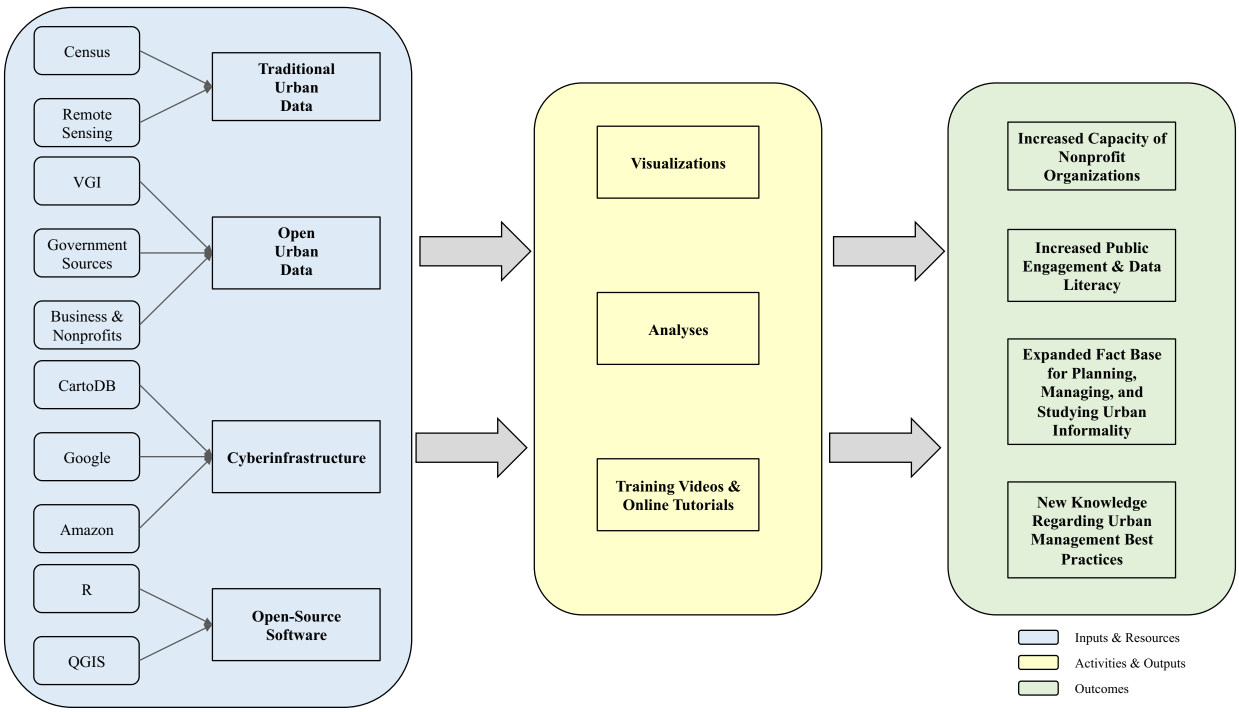

Fig. 3. The logic model.

5. Inputs and resources: components of the system architecture

5.1. Traditional urban data

Between 2005 and 2014 all but five countries in the world committed to conducting a national census (United Nations, 2014b). However, the frequency at which these efforts are undertaken and the level of detail at which information are collected varies significantly from country to country. In India, the Ministry of Home Affairs has conducted a decennial census since 1872 with the most recent data collection effort coming in 2011. Although plans to open data centers where researchers can access microdata were announced in 2012, data below the city or village level are generally not available. Informal settlements data are even less likely to be available, adding an additional layer of complexity for urban managers and researchers. Similarly, in places that are experiencing rapid urbanization and population growth, a decennial census quickly becomes outdated and loses its reliability as a fact base for urban management. As a result of these and other data limitations, a considerable proportion of the research focusing on urban informality has relied on imagery and image-derived datasets acquired through remote sensing (Musakwa & Van Niekerk, 2015).

In addition to the Landsat program which can provide global satellite imagery dating back to the launch of its first sensor in 1972, there are now a host of private (e.g., QuickBird, IKONOS) and national satellites offering data at much higher temporal and spatial resolutions. However, relying extensively in satellite imagery has also influenced the kinds of research questions and practical applications that scholars and urban managers have explored most extensively. The sheer number of high resolution satellite imagery vendors, combined with the opacity of pricing schedules and complexity of licensing requirements, make relying on these resources alone to study urban informality expensive and daunting. For example, in March 2015 the authors acquired two orthorectified QuickBird images covering the K-West ward in Mumbai from an authorized vendor for $1100. The area of interest for both images was 34 square kilometers for a net cost of $16.18 per square kilometer. While the spatial resolution of the imagery (50 cm pan sharpened) is ideal for modeling the expansion of informal settlements (which was the purpose of that project), there are inherent tradeoffs with the extent of the area that can be covered, both from a data storage and cost perspective. Higher spatial resolution means many more pixels to store and analyze and if the goal is longitudinal analysis or change detection, multiple images would need to be purchased.

5.2. Open urban data

The open data movement is radically changing the way researchers and practitioners interact with data and is catalyzing new research initiatives (Kitchin, 2014). In the US, the federal government has established standards and web technologies to facilitate the sharing of information and established “open and machine-readable” data as the default for datasets maintained and released by its agencies (White House, 2014). Cities like Chicago are creating and maintaining online portals where residents, businesses, researchers, and others with an interest in data can access and often visualize or analyze a wide array of data. While the movement toward open urban data in the US was originally “driven by a commitment to transparency and accountability,” the potential for these resources go further into innovation and new forms of public participation and community engagement (Gurin, 2014). The federal government recently launched a website[1] specifically designed to serve as a repository for open data provided not only by federal agencies, but also maintains resources (e.g., data, web and mobile apps, application program interfaces) made available to the public by states, cities, and counties across the nation.

In addition to efforts made by federal, state, and local government, the nonprofit sector has also been instrumental in the proliferation and success of the open data movement. For example, “the MacArthur Foundation, Sunlight Foundation, Code for America and Metro Chicago Information Center provide additional financial and technical assistance to the local government” in Chicago (Kassen, 2013). The potential impact of embracing open data on the work of nonprofit organizations (NPOs) has been summarized as “improved planning by NPOs, more informed decision-making by funders, and researchers and analysts furthering our knowledge of the sector” (Lenczner & Phillips, 2012). The private sector has led the charge in the acquisition and analysis of big data (Manyika et al., 2011), not only as a means of increasing efficiency, but also “for leveraging value from a company’s activities and its

[1]At the time of writing this article, there were over 189,000 datasets available on www.data.gov.

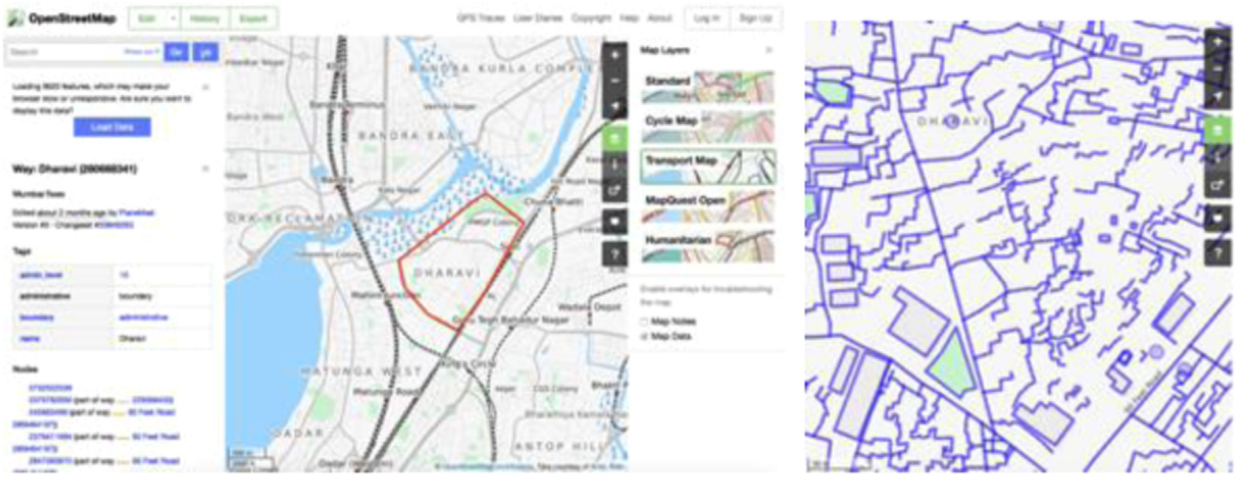

Fig. 4. OSM map of Dharavi slum in Mumbai in two different extents.

relationships with suppliers and consumers" (Kitchin, 2014). Given the profit motive and strategic implications surrounding much of this information, it is not surprising that the business world has been slower to embrace open data (Gurstein, 2011).

Volunteered Geographic Information (VGI) is one of the primary sources of open urban data, which can be used to supplement and fill gaps left by more formal and authoritative data sources like census information or satellite imagery. VGI centers on the notion of "humans as sensors" and rests on the proliferation of technologies (e.g., Web 2.0, global positioning systems, broadband internet connections, and smartphones) that enable the decentralization of geospatial data production (Goodchild, 2007). OpenStreetMap (OSM) is one of the most visible examples of VGI, primarily because it serves a repository for contributed data as well as a platform for mapping and visualization. As alternative databases, these mapping and map-sharing portals parallel proprietary databases such as ESRI and Google and those maintained by the public agencies. Non-profit organizations can use OSM to access basic physical and infrastructure data, such as streets and natural features, for a large number of areas globally. OSM also allows mapping capabilities that grassroots organizations or their constituents may employ to add details to any existing mapped information. While such efforts are common among mapping organizations, the value of such maps and mapping efforts can be broader. For example, in the aftermath of the April 2015 earthquakes in Nepal, many emergency evacuation and relief organizations relied on maps made by the Kathmandu Living Labs to delivery necessary services. In slums in Mumbai, however, there is still little publicly available spatial information on the nature of residents, detailed building footprints, and access to economic and cultural amenities.(Fig. 4)

5.3. Mobile apps

Mobile devices, and smartphones in particular, are remaking both the consumption and production of information in that they allow users to engage in a variety of tasks including mapping, photography, audio and video capture, that "previously required multiple devices, exorbitant resources, and specialized training" (Boone, 2015). The fact that mobile devices tend to come equipped with built-in GPS receivers or can be derived from known positions of cell towers and Wi-Fi access points through triangulation or fingerprinting (Liu, Darabi, Banerjee, & Liu, 2007) means that the data collected by these devices can also be rendered spatially explicit and used for mapping and analysis. While growing number of smartphone users may mitigate the cascading effects of the Digital Divide (Servon & Nelson, 2001), urban managers, planners, nonprofit organizations, and researchers need a better understanding of how target audiences use these mobile technologies in order to fully leverage their potential. One example is Open Data Kit, which is a "free and open source[2] set of tools which help organization author, field, and manage mobile data collection" (Open Data Kit, 2015) and was developed specifically to support mobile data collection in the

[2]Steiniger and Bocher (2009) provide a detailed explanation of what these terms mean.

developing world (Anokwa, Hartung, Brunette, Borriello, & Lerer, 2009; Hartung et al., 2010). The most recent release of Open Data Kit takes advantage of new advances in cloud computing and provides users with a downloadable mobile app (ODK Collect) that can be used to collect locational, text, image, and video information; a web-based form creation and customization application (ODK Build); and a basic database (ODK Aggregate) that supports the storage of blank and completed forms for use with ODK Collect, management of data extracted from submitted forms, simple visualization of those data; and publication to external applications like Google Fusion Tables, OpenStreetMap, or a desktop GIS (Hartung et al., 2010; Brunette et al., 2013). For more advanced users who may want to build a mobile app from scratch or simply have more flexibility in customization, there are several online app design software options including Appsbar[3], which does not charge a fee so long as users agree to allow advertisements into the user interface of the apps they create.

The benefits of these and related technologies has been demonstrated in India where mobile devices and open source GIS were deployed by researchers in collaboration with slum residents succeeded in facilitating interaction and laying the foundation for "effective negotiation and cooperation with local, state and central government agencies" (Livengood & Kunte, 2012, p. 95). The prevalence of mobile phones in India is well known and, according to Nielsen Research (2015), India is poised to be among the largest and fastest growing smartphone markets. These trends promise that smartphone based data collection in India will get easier and, with a greater share of ownership in lower income and slum resident groups, the user data generated by smartphone apps will become more representative.

5.4. Application Program Interface (API) data

An application program interface (API) is a collection of specialized code and functions that are used by software developers, researchers, and advanced users to interact with and manipulate one or more components of an application. A variety of APIs exist that could be useful for urban management including the ubiquitous Google Maps API as well as those maintained by the Indian Ministry of Communications and Information Technology. Registered users of the Indian Open Government Data Platform may request an API key, which allows programmatic access to government maintained data and resources, many of which are presented at a level of detail that is greater than what is presented via official websites. Social media sites such as Facebook, Twitter, and Waze provide access to some of the data they collect via APIs.

5.5. Open source GIS

While Environmental Systems Research Institute of ESRI's ArcGIS remains the most widely used desktop GIS, several open source alternatives (Steiniger & Bocher, 2009) have emerged in recent years that are better able to compete with respect to functionality, robustness, and usability. Of these alternatives, Quantum GIS (QGIS) is among the most popular and well-established and offers server and mobile applications in addition to the desktop application as part of the base installation. QGIS offers a standard plugin that allows more experienced users to leverage the functionality of the older, and less user-friendly Geographic Resources Analysis Support System (GRASS) software suite. A second major component of the QGIS suite is the System for Automated Geoscientific Analyses or SAGA GIS application, which offers basic geoprocessing functionality, along with tools that are well-suited for terrain analysis, image processing, and geostatistics. The convergence across these open source geographic information systems and proprietary GIS—like ArcGIS—in terms of user interface and the types of functionality provided is a testament to the responsiveness of software development communities to the needs of users as well as the realization that geospatial data analysis is fundamental to many of the most urgent policy issues and research questions of the day.

5.6. R

In the early 1990s, a group of academic statisticians began what would become the R project (Ihaka & Gentleman, 1996) as a variant of the S language developed in the 1970s by Bell Labs (Bivand, 2006). In addition to being a

[3]More information is available at http://www.appsbar.com

programming language, R is also a software application that supports data management, visualization, and analysis through a core set of basic functionality and supplemented by "contributed packages addressing the needs of different scientific communities" (Bivand, 2014). The base application as well as contributed packages tailored to specific research and teaching purposes are available through a set of internet mirror sites known as the Comprehensive R Archive Network or CRAN. R is also a community of users and developers with more than 8000 registered users as of May 2014 (Bivand, 2014) and over 7400 packages available for download[4] at the time this article was written. The proliferation of R as a tool for research and teaching has been fueled by the fact that it is available at no charge, users have the ability to "examine, modify, and redistribute the software" (Theußl & Zeileis, 2009), its ability to handle spatial data and to interoperate with GIS, its ability to produce sophisticated and publication-quality graphics, as well as the development of RStudio as an alternative for users who may be less familiar or comfortable with the command line or scripting (Bivand, 2014)[5].

5.7. Cyberinfrastructure

The term cyberinfrastructure first entered the popular discourse in 1998 as part of a White House briefing (Yang, Raskin, Goodchild, & Gahegan, 2010) was later codified by the U.S. National Science Foundation (NSF) in a report released in 2003. This report described cyberinfrastructure as "infrastructure based upon distributed computer, information and communication technology" and noted "if infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy" (Atkins et al., 2003, p. 5). In the context of leveraging open data for managing urban informality, the cyberinfrastructure component (see Fig. 3) could consist of server space and middleware to link the various components of the overall slum information toolkit. The advent of cloud computing and emergence of free and low cost options such as Amazon Web Services and Google Cloud Platform have made it easier for organizations outside government and the private sector, as well as individuals to build and deploy complex applications. This removes the burden of acquiring and maintaining hardware, allowing researchers, urban managers, and nonprofit organizations to focus more directly on data collection, integration, visualization, analysis, and knowledge generation. By coupling basic cyberinfrastructure with open source software or cloud-based mapping and visualization tools like CartoDB, the limitations of traditional urban data with respect to managing informal settlements can be overcome, these areas can be better integrated into formal urban management and planning processes, and new knowledge regarding best practices can be created. These outcomes are consistent with the overarching vision of cyberinfrastructure as a flexible and cross-cutting resource that supports interdisciplinary research and transformative applications of scientific discoveries (Atkins et al., 2003; Yang et al., 2010).

6. Activities and outputs: applications of the system

6.1. Visualization

One example of visualization using open data is the work done by our research team in the K-West ward of Mumbai, a densely populated and diverse region in the West-Central suburbs. The work employed the following free or open source / open data resources: R (http://www.rproject.org/), Open Office (https://www.openoffice.org/), Google Drive/Maps/Earth (www.drive.google.com, www.maps.google.com, www.earth.google.com), QGIS (http://www.qgis.org/en/site/), OSM, www.openstreetmap.org/), and web-portal for creating, visualizing, and sharing maps CartoDB (http://cartodb.com/). The work began by mining the Twitter API using an R-script and a contributed R package called "TwitteR"[6]. The collected information on volume and content of tweets were then analyzed with QGIS alongside other spatial data, such as road network and major activity centers downloaded from OSM. (Fig. 5)

We hypothesized that simply organizing the volume of tweets by time of day would allow us to distinguish commercial areas (higher activity during the workday) from residential (higher activity in the evening and early morning). A visit by

[4] See the Comprehensive R Archive Network (CRAN) website for details https://cran.r-project.org.

[5] A number of open-source tools are now available that are designed for more specialized environments. For example, Epi Info™ is a tool designed for the public health community. It is a package of easy to use tools that supports data collection, complex statistical analysis and mapping functionalities (http://wwwn.cdc.gov/epiinfo/).

[6] Specific details on how to mine Twitter using R can be found at https://heuristically.wordpress.com/2011/04/08/text-data-mining-twitter-r/, last accessed October 28, 2015.

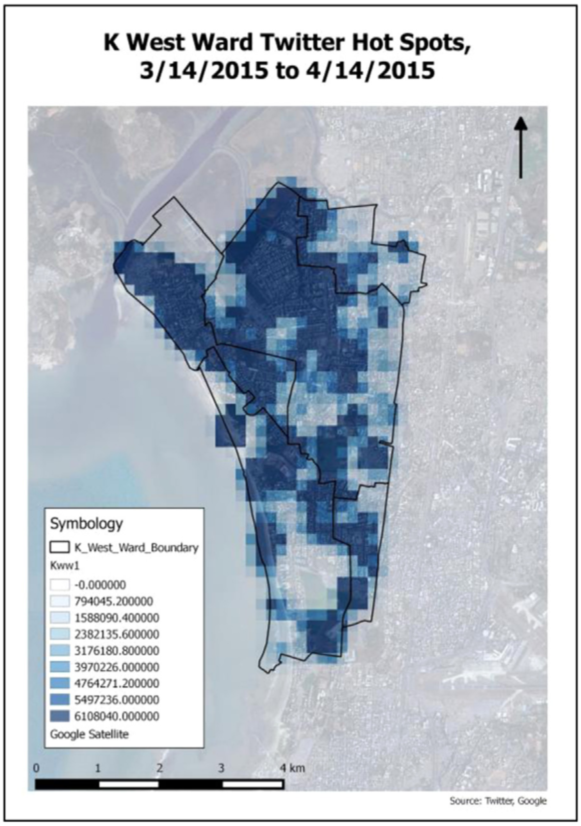

Fig. 5. Twitter hotspots map for K-West Ward of Mumbai created in QGIS [by Elizabeth Bastian].

members of our research team to these sites in Mumbai and crosschecking other areas using aerial imagery on Google Maps confirmed our expectations. Going by the assumption that tweets are not representative of the larger population, and less popular among lower income groups, we also expected to see relatively little Twitter activity inside informal settlements. This was true for some informal settlements and not others. On visiting the slum area with high Twitter activity, we found a private health clinic alongside a community area. Finally, we can also analyze the extracted tweets qualitatively for their content and linguistic characteristics. As Mumbai slums are home to many migrant communities from around the country, organizing these datasets by frequency of specific languages used in tweets and the ideas and concerns expressed in those tweets can provide some information about the places, their residents, and surrounding infrastructure. Finally, a CartoDB environment was created to visualize this dataset and with added capabilities of querying the underlying data.(Fig. 6)

This example highlights how many of the components outlined in our framework (Fig. 3) can be combined to better understand access to technology and potentially, cultural diversity as well as mobility within informal areas. For example, R code that allows users to automatically detect the language of a Tweet has been developed (Charpentier, 2015) and could be used to study the distribution of ethnic and cultural groups in the city as well as changes in these patterns over time. Understanding the geography of local vernaculars could be useful for communication during an emergency or as a means of targeting public engagement efforts. Data mining algorithms exist that identify and order tweets from the same user over time, making it possible to capture and study both the "regularities and variability" of individual travel behavior in areas where formal surveys or models do not exist (Huang & Wong, 2015).

6.2. Analysis

Open urban data also allows us to conduct a variety of analysis for planning and urban management. A tool similar to OpenDataKit used in slums in India is EPI Info. A recent project by a member of the research team used EPI Info

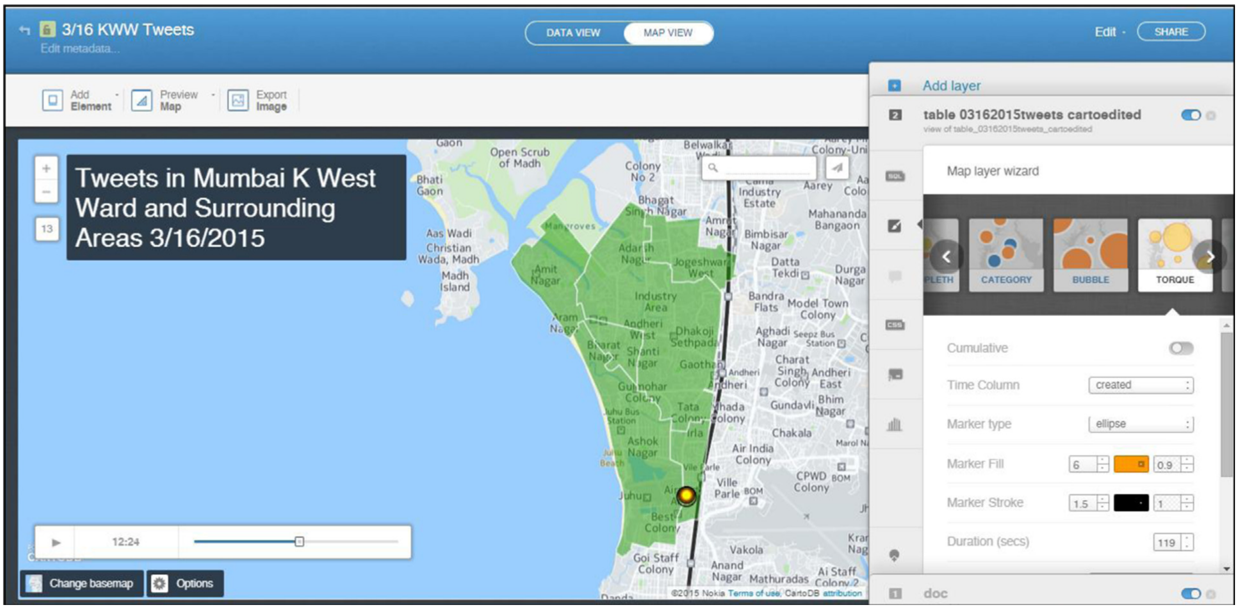

Fig. 6. CartoDB interface allows users to add additional information and query the underlying data [Prepared by Elizabeth Bastian].

and community-based surveyors with little experience to conduct a geo-coded household survey in slums. The survey was used to understand socioeconomic characteristics of slums that grew at the urban periphery and to extend this knowledge to assess the patterns of slum growth. EPI Info generates some basic analysis automatically but with a little financial resources and time, the researchers were also able to conduct more rigorous data validation, statistical analysis and mapping. The current version of the tool also provides interface for mobile devices. Mapping the surveyed information on age of units, for example, shows that the settlement started about 25 years ago with a cluster of few houses near publically available land and grew northwards. A second and newer cluster started forming in the south attributed to migration.(Figs. 7 and 8)

Another example of analysis using open data is a land use change model our research team built for Mumbai Metropolitan Region (MMR) using free and open data. Land use change models operationalize the interdependencies between components of the built environment such as roads, employment centers, residential neighborhoods, green space, municipal boundaries, etc. and help explore their implications for the future. We adopted the SLEUTH model (Clarke, Hoppen, & Gaydos, 1997), a free cellular automata (CA) model that uses data on areas resistant to urban development (Slope), land use and land cover (Land use), areas where urbanization cannot occur (Excluded), urban extent (Urban), transportation (Transportation), and topography (Hillshade). The data were collected from the following sources, all open and free. Urban extents, land use, and land cover data for multiple years were derived from Landsat orthoimagery available through the Global Land Survey (GLS). Road data layers were created from OpenStreetMap Metro Extracts and crosschecked with MMR Comprehensive Transportation Plan maps. The exclusion layer that includes water bodies, parks, forest, sport arenas, airport, and train stations was created from OSM data. The slope layer was derived from the Global Digital Elevation Model (GDEM) built on NASA's Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). The SLEUTH model was compiled, calibrated, and executed under the Cygwin environment, a Windows- based UNIX emulator. An output of the model showing areas of high growth probability in the region is presented 8.

Land use change models can serve a variety of functions, such as testing the future impact of present day policies, identifying areas under high development pressure, and examine equity implications of ongoing development trends. Our model was able to identify high growth pressures in areas marked for future coastal preservation as well as areas where infrastructure is already overburdened. This has implications for policymakers who may wish to discourage development in such areas and encourage them in others. While our model included both formal and informal developments, SLEUTH model is unable to distinguish between different types of developed residential areas. In order to correct that, we are currently using open data to develop a land use change model specifically for informal settlements.

6.3. Training and tutorials

While some urban datasets and techniques have been around for a while, a number of barriers in the past have inhibited NGOs and grassroots organizations from using them. This has been attributed to the lack of technical capacity of such organizations in obtaining and using data (Elwood, 2008), lack of knowledge about data resources (Ghose, 2001), financial resources, hardware and network capacity (Barndt, 1998), and even political connections (Elwood, 2008). Many of these barriers continue to be relevant. But, fortunately, resources are now emerging that provide free and accessible training to match the interests of a variety of stakeholders.

Among the most commonly used learning tools are YouTube videos. There are a host of videos providing tutorials on OpenStreetMaps, R, and QGIS. In addition, there are a number of tutorials specializing in the use of smartphone apps such as creating surveys using Google Forms, and OpenDataKit. Many of the smartphone apps allow tagging

Fig. 7. Mapping slum growth using data collected in EPI Info (*Note: The image was not created within Epi Info™).*

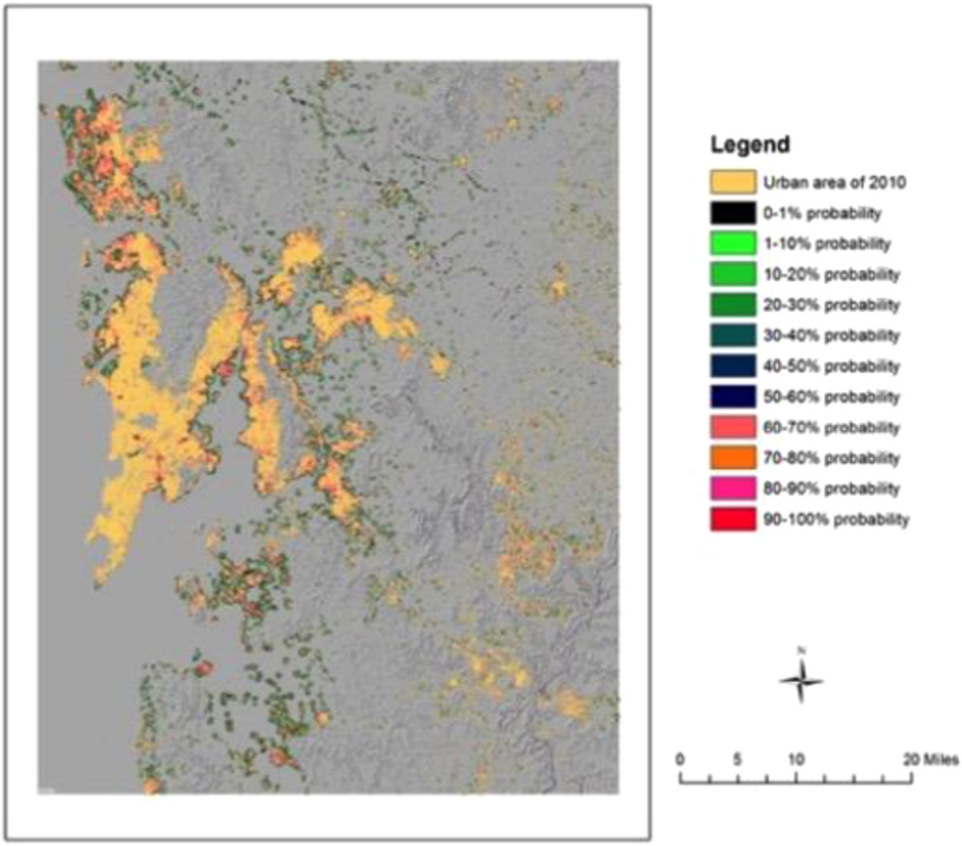

Fig. 8. Growth probability map for Mumbai metropolitan region for 2030.

unique GPS identifiers or locations of data collections and a number of tools even allow cross platform integration such as OpenStreetMap and OpenDataKit. Most of these tools also come with support information[7].

A number of platforms such as OSM also facilitate map development workshops or Mapathons. Kathmandu Living Labs, a non-governmental organization in Nepal mentioned earlier in the paper, works with OSM in hosting such events directly in the communities and provide access to resources to others in using open data. These training and mapping events empower the community members in a number of ways, including advancing local governance and citizen engagement and mapping previously unmapped critical infrastructures.

7. Outcomes and conclusion: the promise of open urban data

In the last box of our logic model, we identify some possible outcomes from open urban data literacy and usage, including increased capacity of nonprofit organizations, increased public engagement, and new knowledge, among other positive outcomes. A number of recent and ongoing activities with regard to informal settlements in Mumbai underscore these potential values and show how different components of our framework can come together in real life applications.

The first example speaks to how the lack of consistent datasets *(inputs)*, especially with regard to informal settlements, can influence in a planning process *(activity)*, namely the ongoing Development Plan update for Greater Mumbai, and ultimately hinder outcomes for both formal and informal development. The Municipal Commission for Greater Mumbai (MCGM) is undertaking a statutory development plan update for the first time since 1994[8], and the role of urban data has been of much contention. As noted before, many (though not all) slum areas in Mumbai fall

[7]See, for example: http://learnosm.org/en/beginner/introduction/.

[8]The planning process started in 2010 and has dragged past its initial expected completion date due to controversy.

Fig. 9. A map of Gautam Nagar informal settlement in Mumbai created by child advocates with assistance from Humara Bachpan (Photo Credits: Elizabeth Bastian and Andrew McMillan).

under the jurisdiction of Slum Rehabilitation Agency. While MCGM's planning studies have included a survey of slum areas and even called for setting up a “slum information system”, the actual draft plan pays far less attention to slums and has been accused of incorrectly representing boundaries of slum areas. Indeed, many of the plan's maps show grayed out zones in areas where hundred of thousands of slum residents live. The acute shortage of affordable housing in the Mumbai, the fact that more than two-fifths of its population lives in slums, the strong emphasis on development capacity in the plan, makes the lack of adequate consideration of slums in the planning process especially remarkable. Other data controversies have included mapping of coastal resource zones and parks. These and other issues seemingly contributed to strong opposition to the plan by many civic groups and led the state government to issue a directive under the state's “Regional and Town Planning Act 1966” to examine and re-publish the plan at a later date.

While the above problem is not entirely urban data related, we believe that the proposed slum information system, and more broadly, an urban information system can be a strong addition to the planning toolkit and help reduce uncertainty over factual basis of spatial information. A slum information system can build on existing efforts, such as the Municipal Corporation of Greater Mumbai’s ‘Voice of Citizens’ portal (see http://www.voiceofcitizen.com/). Voice of Citizens allows citizens to report public service related complaints, such as potholes, directly to the public works department. Users may register their complaints through a web browser or a mobile application and even submit geo tagged photographs. While still in its infancy, these approaches can serve as the basis for tracking service responses in both formal and informal settlements and instill transparency in service delivery. Moreover, similar to the Twitter example described earlier, the data generated can help us better understand access to public services and maintenance across areas of different types. If there are barriers to municipal agencies in setting up completely open systems, local institutions such as the Indian Institute of Technology in collaboration with local civic and non-governmental organizations may assemble the basic infrastructure for such a system. Such examples of university based data centers are common elsewhere and they can help support planning decisions as well as academic research.

The second example is more bottom-up and speaks directly to the exciting possibilities of incorporating more digital resources for organizations working in informal settlements. Some members of our research team recently had the opportunity to spend a day with the coordinators of Humara Bachpan (translated Our Childhood) and “child advocates" of the Gautam Nagar community, an old informal settlement next to the new international airport in Mumbai. Humara Bachpan runs campaigns across India, including in Gautam Nagar, for a "safe and healthy environment for children living in urban poverty". A common tool employed by this group is hands-on drawing, where the child advocates draw maps as a way to communicate the present state of their community as well as their aspirations to those in power. We found this to be highly inspiring and impactful work (Fig. 9). But at the time of our visit none of this work was in digital format. Nor was there a way to connect these maps to other aspects of their community or plans of surrounding areas.

No doubt a hand-drawn map by a child has a unique value that cannot be translated to digital maps. Neither can all the controversies of a citywide development plan be removed solely by better data. Nevertheless, both processes can use better data, and the ability to connect one dataset to another and conduct additional analysis. And using open urban data and software can be especially valuable in informal settlements because such communities have greater resource-constraints. Open urban data and software are improving constantly and the increasing popularity of smartphones and vehicular devices means more citizens are effectively becoming sensors. If the risk to privacy can be managed, the opportunities seem immense. Indeed, these new technologies – and the opportunity for data-driven decision making they create – can be as effective in informal settlements as they are purported to be in future smart cities. In that sense, the framework presented in this paper enhances and broadens the Smart Cities discourse by making urban management more efficient and helps to more fully realize the transformative potential of technology.

References

Abbott, J. (2001). Use of spatial data to support the integration of informal settlements into the formal city. *International Journal of Applied Earth Observation and Geoinformation*, *3*(3), 267–277.

Anokwa, Y., Hartung, C., Brunette, W., Borriello, G., & Lerer, A. (2009). Open source data collection in the developing world. *Computer*, *42*(10), 97–99.

Arputham, J., & Patel, S. (2010). Recent developments in plans for Dharavi and for the airport slums in Mumbai. *Environment and Urbanization*, *22*(2), 501–504.

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., et al. (2003). Revolutionizing science and engineering through cyberinfrastructure: *Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. Retrieved from ⟨https://www.nsf.gov/cise/sci/reports/atkins.pdf⟩.

Banerjee, B. (2002). Security of tenure in Indian cities. In: A. Durand-Lasserve, & L. Royston (Eds.), *Holding their ground: secure land tenure for the urban poor in developing countries*, 37–58). London, UK: Earthscan.

Bardhan, R., Sarkar, S., Jana, A., & Velaga, N. R. (2015). Mumbai slums since independence: evaluating the policy outcomes. *Habitat International*, *50*, 1–11.

Barndt, M. (1998). Public participation GIS—barriers to implementation. *Cartography and Geographic Information Systems*, *25*(2), 105–112.

Batty, M. (2008). The size, scale, and shape of cities. *Science*, *319*(5864), 769–771, http://dx.doi.org/10.1126/science.1151419.

Baud, I., Scott, D., Pfeffer, K., Sydenstricker-Neto, J., & Denis, E. (2014). Digital and spatial knowledge management in urban governance: emerging issues in India, Brazil, South Africa, and Peru. *Habitat International*, *44*, 501–509.

Bivand, R. (2006). Implementing spatial data analysis software tools in R. *Geographical Analysis*, *38*(1), 23–40.

Bivand, R. (2014). The R Development Process: status and prospects. Retrieved from ⟨http://estymator.ue.poznan.pl/blog/pazur/vol7/RogerBivand.pdf⟩.

Boone, K. (2015). Disembodied voices, embodied places: mobile technology, enabling discourse, and interpreting place. *Landscape and Urban Planning*, *142*, 235–242.

Brunette, W., Sundt, M., Dell, N., Chaudhri, R., Breit, N., Borriello, G. (2013). Open data kit 2.0: expanding and refining information services for developing regions. In: *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, Association for Computing Machinery (p. 10).

Budhathoki, N. R., Bruce, B., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*, *72*(3–4), 149–160.

Chakraborty, A., Wilson, B., & Bin Kashem, S. (2015). The pitfalls of regional delineations in land use modeling: implications for Mumbai region and its planners. *Cities*, *45*, 91–103.

Charpentier, A. (2015). "Automatic detection of the language of a tweet." Retrieved from ⟨http://freakonometrics.hypotheses.org/18690⟩.

Clarke, G. (1991). Urban management in developing countries: a critical role. *Cities*, *8*(2), 93–107.

Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, *24*, 247–261.

Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Stefanidis, A., Smith, D., & Lamprianidis, G. (2014). Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 1–22.

Dagdeviren, H., Robertson, S. A. (2009). Access to water in the slums of the developing world (No. 57). Working Paper, International Policy Centre for Inclusive Growth.

Das, A. K., & Takahashi, L. M. (2009). Evolving institutional arrangements, scaling up, and sustainability emerging issues in participatory slum upgrading in Ahmedabad, India. *Journal of Planning Education and Research*, *29*(2), 213–232.

Das, P. K. (2011). Mumbai's Slums Map. Accessed 23.11.15, ⟨http://www.pkdas.com/maps/3-Mumbai's-Slums-Map.pdf⟩.

Das, P. K. (2012). Open Mumbai: Re-envisioning the City and its Open Spaces. Mumbai, India: Mumbai Waterfronts Centre and P.K. Das & Associates.

Donovan, K. P. (2012). Seeing like a slum: towards open, deliberative development. *Georgetown Journal of International Affairs*, *13*, 97.

Doshi, S. (2013). The politics of the evicted: redevelopment, subjectivity, and difference in Mumbai's slum frontier. *Antipode*, *45*(4), 844–865.

Elwood, S. (2008). Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*, *22*(1), 71–90.

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: effective use of theories of change and logic models*. San Francisco, CA: John Wiley & Sons.

Ghose, R. (2001). Use of information technology for community empowerment: transforming geographic information systems into community information systems. *Transactions in GIS*, *5*(2), 141–163.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *Geo Journal*, *69*(4), 211–221.

Goel, V., Facebooks Internet for All Is a Tough Sell in India. New York Times, ⟨http://www.nytimes.com/2015/10/26/technology/facebook-meets-skepticism-in-bid-to-expand-internet-in-india.html?smid=fb-share⟩, Accessed 30.10.15.

Government of India. (2010). Report of the Committee on Slum Statistics/Census. New Delhi, India: Ministry of Housing and Urban Poverty Alleviation, National Buildings Organisation. ⟨http://nbo.nic.in/Images/PDF/REPORT_OF_SLUM_COMMITTEE.pdf⟩, Accessed 23.11.15.

Gramlich, E. M. (1994). Infrastructure investment: a review essay. *Journal of Economic Literature*, 1176–1196.

Gurin, J. (2014). "How open data is transforming city life." Forbes, 12 September 2014. Retrieved from ⟨http://www.forbes.com/sites/techonomy/2014/09/12/how-open-data-is-transforming-city-life/⟩.

Gurstein, M. B. (2011). Open data: empowering the empowered or effective data use for everyone?. *First Monday*, *16*, 2.

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., Borriello, G. (2010). Open data kit: Tools to build information services for developing regions. In: *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development* (p. 18).

Hopkins, L. D. (2001). *Urban development: the logic of making plans*. Washington, DC: Island Press.

Huang, Q., & Wong, D. W. (2015). Modeling and visualizing regular human mobility patterns with uncertainty: an example using Twitter data. *Annals of the Association of American Geographers*, *105*(6), 1179–1197.

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

Kassen, M. (2013). A promising phenomenon of open data: a case study of the Chicago open data project. *Government Information Quarterly*, *30*(4), 508–513.

Kellogg Foundation, W. K. (2004). Using logic models to bring together planning, evaluation, and action: Logic model development guide. Battle Creek, MI: W.K. Kellogg Foundation.

Kitchin, R. (2014). *The data revolution: big data, open data, data infrastructures and their consequences*. London, UK: Sage Publications.

Klosterman, R. E. (1994). Large-scale urban models retrospect and prospect. *Journal of the American Planning Association*, *60*(1), 3–6.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J. F., & Bertollini, R. (1999). *Disease mapping and risk assessment for public health*. New York, NY: John Wiley & Sons.

Lee, J. (2007). A three-dimensional navigable data model to support emergency response in microspatial built-environments. *Annals of the Association of American Geographers*, *97*(3), 512–529.

Lenczner, M., & Phillips, S. (2012). From stories to evidence: how mining data can promote innovation in the nonprofit sector. *Technology Innovation Management Review*, *2*, 7.

Liu, H., Darabi, H., Banerjee, P., & Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *37*(6), 1067–1080.

Liu, X., Song, Y., Wu, K., Wang, J., Li, D., & Long, Y. (2015). Understanding urban China with open data. *Cities*, *47*, 53–61.

Livengood, A., & Kunte, K. (2012). Enabling participatory planning with GIS: a case study of settlement mapping in Cuttack, India. *Environment and Urbanization*, *24*(1), 77–97.

McFarlane, C., Desai, R., & Graham, S. (2014). Informal urban sanitation: Everyday life, poverty, and comparison. *Annals of the Association of American Geographers*, *104*(5), 989–1011.

Maddock, A., & Du Plessis, M. A. (1999). Can species data only be appropriately used to conserve biodiversity?. *Biodiversity Conservation*, *8*(5), 603–615.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: the next frontier for innovation, competition, and productivity*. Washington, DC: McKinsey Global Institute.

McFarlane, C. (2012). The entrepreneurial slum: civil society, mobility and the co-production of urban development. *Urban Studies*, *49*(13), 2795–2816.

MCGM. (2014). Development Plan for Greater Mumbai 2014-2034: Preparatory Studies. Retrieved from ⟨http://mcgm.gov.in/irj/portal/anonymous/qlddevplan⟩ Accessed 01.09.15.

Millar, A., Simeone, R., & Carnevale, J. (2001). Logic models: a systems tool for performance management. *Evaluation and Program Planning*, *24*, 73–81.

Musakwa, W., & Van Niekerk, A. (2015). Earth observation for sustainable urban planning in developing countries needs, trends, and future directions. *Journal of Planning Literature*, *30*(2), 149–160.

Nielsen Research, Smartphones keep users in India plugged in, ⟨http://www.nielsen.com/us/en/insights/news/2013/smartphones-keep-users-in-india-plugged-in.html⟩, Accessed 30.10.15.

Nolan, L. B. (2015). Slum definitions in urban india: implications for the measurement of health inequalities. *Population and Development Review*, *41*(1), 59–84.

Open Data Kit. (2015). About Open Data Kit. Retrieved from ⟨https://opendatakit.org/about⟩.

Patwa, S. (2013). Slums ignored in BMC's new development plan. November 25, 2013. Indian Express. Retrieved from ⟨http://archive.indianexpress.com/news/slums-ignored-in-bmc-s-new-development-plan/1199121/⟩.

Pinto, R. (2015). Include affordable housing, informal sectors in draft DP. April 28, 2015. The Times of India. Retrieved from ⟨http://timesofindia.indiatimes.com/city/mumbai/Include-affordable-housing-informal-sectors-in-draft-DP/articleshow/47064435.cms⟩.

Rambaldi, G., Kyem, P. A.K., McCall, M., & Weiner, D. (2006). Participatory spatial information management and communication in developing countries. *The Electronic Journal of Information Systems in Developing Countries*, *25*.

Roy, A. (2009). Why India cannot plan its cities: informality, insurgence and the idiom of urbanization. *Planning Theory*, *8*(1), 76–87.

Sclar, E. D., Garau, P., & Carolini, G. (2005). The 21st century health challenge of slums and cities. *The Lancet*, *365*(9462), 901–903.

Sen, S. (1999). Some aspects of State–NGO relationships in India in the post-independence era. *Development and Change*, *30*(2), 327–355.

Servon, L. J., & Nelson, M. K. (2001). Community technology centers: narrowing the digital divide in low-income, urban communities. *Journal of Urban Affairs*, *23*(3 & 4), 279–290.

Steiniger, S., & Bocher, E. (2009). An overview on current free and open source desktop GIS developments. *International Journal of Geographical Information Science*, *23*(10), 1345–1370.

Theußl, S., & Zeileis, S. (2009). A. collaborative software development using R-Forge. *The R Journal*, *1*(1), 9–14.

Turley, R., Saith, R., Bhan, N., Doyle, J., Jones, K., & Waters, E. (2013). Slum upgrading review: methodological challenges that arise in systematic reviews of complex interventions. *Journal of Public Health*, *35*(1), 171–175.

UN-Habitat, E. (2003). *The challenge of slums: global report on human settlements 2003*. London, UK: Earthscan.

United Nations. (2014a). World urbanization prospects: The 2014 revision, highlights (ST/ESA/SER.A/352). New York, NY: United Nations, Department of Economic and Social Affairs, Population Division.

United Nations. (2014b). 2010 World Population and Housing Census Programme: Q & A. New York, NY: United Nations Statistics Division. Retrieved from ⟨http://unstats.un.org/unsd/demographic/sources/census/2010_phc/QA.htm⟩.

Venkatraman, T. (2015). Draft Development Plan: 35,000 letters processed so far. May 22, 2015. The Indian Express. Retrieved from ⟨http://indianexpress.com/article/cities/mumbai/draft-development-plan-35000-letters-processed-so-far/⟩.

Weinstein, L. (2014). One-man handled': fragmented power and political entrepreneurship in globalizing Mumbai. *International Journal of Urban and Regional Research*, *38*(1), 14–35.

Werlin, H. (1999). The slum upgrading myth. *Urban Studies*, *36*(9), 1523–1534.

White House (2014) Open Data Action Plan, May 9 2014, http://www.whitehouse.gov/sites/default/files/microsites/ostp/us_open_data_action_plan.pdf, Accessed: December 1, 2015.

White, R., & Engelen, G. (2000). High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, *24*(5), 383–400.

Yang, C., Raskin, R., Goodchild, M., & Gahegan, M. (2010). Geospatial cyberinfrastructure: past, present and future. *Computers, Environment and Urban Systems*, *34*(4), 264–277.

Zérah, M. H. (2008). Splintering urbanism in Mumbai: contrasting trends in a multilayered society. *Geoforum*, *39*(6), 1922–1932.

Zetter, R., & De Souza, F. A. (2000). Understanding processes of informal housing: appropriate methodological tools for a sensitive research area. *International Planning Studies*, *5*(2), 149–164.