Arguments for and Field Experiments in Democratizing Digital Data Collection – The Case of Flocktracker

By

Daniel Heriberto Palencia Arreola

Civil Engineer
Universidad Nacional Autónoma de México
Mexico City, Mexico (2012)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of
Master in City Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2019

Author ____
Department of Urban Studies and Planning
February 8, 2019

Certified by ____
Associate Professor P. Christopher Zegras
Department of Urban Studies and Planning
Thesis Supervisor

Accepted by ____
Associate Professor P. Christopher Zegras
Chair, MCP Committee
Department of Urban Studies and Planning

Arguments for and Field Experiments in Democratizing Digital Data Collection – The Case of Flocktracker

By

Daniel Heriberto Palencia Arreola

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of
Master in City Planning

Abstract

Data is becoming increasingly relevant to urban planning, serving as a key input for many conceptions of a "smart city." However, most urban data generation results from top-down processes, driven by government agencies or large companies. This provides limited opportunities for citizens to participate in the ideation and creation of the data used to ultimately gain insights into, and make decisions about, their communities. Digital community data collection can give more inputs to city planners and decision makers while also empowering communities. This thesis derives arguments from the literature about why it would be helpful to have more participation from citizens in data generation and examines digital community mapping as a potential niche for the democratization of digital data collection.

In this thesis, I examine one specific digital data collection technology, Flocktracker, a smartphone-based tool developed to allow users with no technical background to setup and generate their own data collection projects. I define a model of how digital community data collection could be "democratized" with the use of Flocktracker. The model envisions a process in which "seed" projects lead to a spreading of Flocktracker's use across the sociotechnical landscape, eventually producing self-sustaining networks of data collectors in a community. To test the model, the experimental part of this research examines four different experiments using Flocktracker: one in Tlalnepantla, Mexico and three in Surakarta, Indonesia. These experiments are treated as "seed" projects in the democratization model and were setup in partnership with local NGOs. The experiments were designed to help understand whether citizen participation in digital community mapping events might affect their perceptions about open data and the role of participation in community data collection and whether this participation entices them to create other community datasets on their own, thus starting the democratization process.

The results from the experiments reveal the difficulties in motivating community volunteers to participate in technology-based field data collection. While Flocktracker proved easy enough for the partner organizations to create data collection projects, the technology alone does not guarantee participation. The envisioned "democratization" model could not be validated. Each of the experiments had relatively low levels of participation in the community events that were organized. This low participation, in turn, led to inconclusive findings regarding the effects of community mapping on participants' perceptions and on the organizations themselves. Nonetheless, numerous insights emerge, providing lessons for the technology and how it might be better used in the future to improve digital community mapping events.

Thesis Supervisor: P. Christopher Zegras

Title: Associate Professor

Acknowledgements

To everyone that keeps teaching me who I want to be and to everyone who reminds me who I do not want to be.

To my advisor, professor Chris Zegras, for all the support given to me.

To my reader, professor Sarah Williams, for your insights and support.

To Jason McKnight, Sandra Wellford, Sylvia Hiestand, MIT's ISO and DUSP, for not letting me slip through the cracks of the system.

To the Flocktracker team: Ali J Al-Sammarraie, Scott Gilman, SriVardhini Suresh and Jenny Wang. Thank you for believing in what we do, for all the sleepless nights, all the empowering discussions and all the personal growth you catalyzed. It has always been a pleasure to work with you.

To Kota Kita and Planeación y Desarrollo, especially Hasanatun Nisa Thamrin, Titis Efrindu Bawono, Ilse Jiménez Molina and Rossio Jiménez Molina. This research would have been impossible without your input and hard work.

To my mother, Luz María del Carmen Arreola Castellanos, my sister, Luz del Carmen Palencia Arreola and my life partner, Zi Xin Wong, for always believing in me.

To Cressica Brazier, Chaewon Ahn, Erina Keefe, Haleemah Qureshi, Ricardo Sánchez, Akemi Sato Matsumoto, Antonio Cano Méndez, Fitse Anley, Adam Hasz and Tia Vice, for all the meals, climbing walls, bike rides, coworking times and the relaxing times too.

To Albert Ching, Kuan Butts, Stephen Kennedy and Arturo Cadena. Every time we talk, I keep thinking about it for months. I am grateful to know you.

To my friends around the world and my cohort at DUSP, for keeping my mind interested, my body active and my heart warm.

To CONACYT and the José Miguel Bejos Fellowship for believing in my work and supporting my studies at MIT.

To the Deshpande Center for Innovation and the Singapore-MIT Alliance for Research and Technology for supporting Flocktracker and my research.

Thank you all.

Contents

Index of Figures and Tables

1 Introduction

Data have always played a role in urban planning. These data can be collected by government authorities (e.g., census, travel surveys), private companies, and/or citizens themselves. These data can be about the population of people and firms, about specific incidents, about environmental conditions, and so forth.

As technological advances further integrate themselves into daily human activities, they give people access to relatively cheap smartphones that come with highly accurate sensors (GPS, microphones, accelerometers, barometers, etc.) and customizable software. This technology gives people increased opportunities to capture “digital traces” of their lives. Concurrent to this increased interest in tracking physical world behavior with digital logs, there has been a movement towards “open data” by governments, at all scales. The proliferation of transparent public data has also spurred (and been spurred by) increasing citizen activism related to “open data.” In this context, digital technologies hold some promise to help “everyday” citizens play a more proactive role in collecting data on topics of interest to them and converting those into meaningful information.

Although enthusiasm towards and discussions about citizen and government activity in this area have become common, implementation of such methods, particularly with regards to new data collection efforts, remains difficult. That is, the promise of digital technology in civic activity (such as a community mapping project) faces often insurmountably high startup friction. For example, minor variations in use cases can necessitate ad-hoc technology development.

Unfortunately, such development requirements can be far too complex or costly for a small community group. Moreover, these efforts often prove to be quite difficult for government entities as well. Issues from procurement procedures to lack of in-house familiarity with digital solution development often result in costly endeavors with subpar results.

Considering these challenges to implementing a somewhat customized application to assist in survey and mapping efforts, private companies have developed business models around providing tools to assist in enabling users to assemble common survey components and mapping-related features into custom survey interfaces. These companies and products (such as Qualtrics, Survey123 or Fulcrum) are becoming more commonplace for field data collection but are not commonly used in community data collection. Online surveys using tools like Google Forms or Survey Monkey are becoming ubiquitous for several fields where spatial categorization and analysis are not required but hard to setup for in-field data collection contexts as their design targets online data collection. Furthermore, most systems are designed primarily around the notion of a consistent internet connection and, as such, they do not work well for field contexts where a reliable web connection is not available, like those encountered in rural areas, difficult terrains and often in developing countries.

Considering the current availability of tools directed specifically at civic engagement and urban mapping projects, service gaps in urban related software still exist. Specifically, a tool that enables nontechnical users (e.g. community activists, advocacy groups) to assemble semi-customizable mapping and surveying interfaces in an intuitive and financially affordable manner could have direct benefits to urban planning practice.

This thesis examines some of these prospects, in light of a specific digital field data collection tool called Flocktracker. Flocktracker is a data collection platform that capitalizes on the increasing availability of internet and smartphones which offer the opportunity to create better, more technified, ways to collect community datasets by taking advantage of the flexibility of software, the safety and reliability of cloud computing, and smartphones' multiple sensors (including GPS and accelerometers) and media (e.g., audio recordings, pictures and video). Flocktracker originated as a purpose-built tool for researching public transport in developing countries, but it has evolved through iterative prototyping to become a flexible in-field data collection tool that allows users to customize their data collection projects to the specific needs of each site and community. Flocktracker aims to be inclusive across different levels of technology literacy because its usage does not require prior training in computer science or geographic information systems (GIS). As part of the research conducted for this thesis, I further developed the Flocktracker platform, and I also conceptualized and operationalized a model to observe the potential "democratization" of a tool like Flocktracker to see whether interconnected communities of data collectors might emerge from the process. This study also aims to help participants in the digital community mapping events conducted to learn more about the role of data collection and participation in their communities and to evaluate the potential partnerships that could make the "democratization" process happen.

1.1 Research questions

In this thesis I aim to answer three basic questions regarding the potential impacts of a digital data collection tool such as Flocktracker and potential strategies to spread its use over communities:

- Can Flocktracker and its deployment model serve the needs of local community organizations and their communities when gathering in-field data?
- Does participating in a digitally enabled community data collection process change participants' perceptions about the role of data and community participation?
- Can an initial digital community mapping experience motivate local community members to undertake their own data collection activities (i.e., "democratize" data collection or make it more widely practiced)?

1.2 Objectives

Through the research undertaken in this thesis, I aimed:

- To work with NGOs to develop community based participatory mapping projects and learn about the challenges they face when setting up and launching projects like this one;
- To identify what kinds of strategies and partnerships would be beneficial to spread the usage of digital data collection in a community;

- To characterize different development stages of a data collection community, from a small niche of users to a democratized tool that is widely used, and develop metrics to observe and measure them;
- To understand what stage of democratization was achieved by the projects developed by taking advantage of metadata generated by the Flocktracker platform and other data sources; and,
- To develop recommendations to conduct community-based digital mapping projects with the overall goal of making these processes easier and more widespread.

1.3 Thesis organization

The remainder of this thesis consists of:

- Chapter 2 – Providing a literature review describing the role of data in urban planning and the rationale for creating new technology-driven tools for bottom-up data collection;
- Chapter 3 – Describing the Flocktracker platform, its design drivers, development process, high level technical components and a model of its potential spread in a community;
- Chapter 4 – Describing the experimental setup for the community mapping projects of this thesis;
- Chapter 5 – Describing the development and outcomes of the four experiment cases, three in Indonesia and one in Mexico; and,

- Chapter 6 – provides conclusions, reflections and recommendations for further digital community mapping projects.

2 Background

In this Chapter, I first outline current urban planning practices in relation to two of its most technologically driven and often interrelated areas: urban data and smart cities. The focus will be on the limited participatory spaces offered by the smart city paradigm and modern urban data practices, the difficulties experienced by urban data-driven decision-making, and the negative consequences already experienced by communities living in areas with little to no data available. This Chapter also highlights some insights from the literature about how participation-driven modern urban technologies can be imagined. Thereafter, I describe potential benefits of the democratization of bottom-up digitally enabled data collection technologies for urban planning and societies. I then examine digital community mapping as an opportunity for citizens to generate their own data, tell their own stories, and participate in planning processes. Then, I introduce the concept of strategic niche management that seeks to systematize means to introduce new technologies to the sociotechnical landscape. The niche use of digital community mapping by NGOs and their communities will be treated in this thesis as the experimental base to understand possibilities for democratizing data collection in the sociotechnical landscape of developing world cities. Finally, I give an overview of Flocktracker as a tool developed to respond to and complement current urban data collection approaches while aiming to facilitate community-led efforts.

2.1 Urban data practice

Urban data is a term applied to information that characterizes human settlements, including socioeconomics, demographics, infrastructure, environmental conditions and historic trends in growth and change (Batty M. , 2008) (Crooks, Pfoser, Jenkins, & Croitoru, 2015). Making decisions based on analyses of data rather than purely on intuition is called data-driven decision-making (Provost & Fawcett, 2013). High quality data that serve as a fact base are a key component for decision-making across a wide variety of sectors and urban scales (White & Engelen, 2000) (Klosterman, 1994). With a data-driven understanding of existing and future urban conditions, urban professionals like architects, developers and planners can better collaborate and make more informed decisions (Ferreira, et al., 2015). Urban data-driven decision-making can benefit diverse areas such as education policy (Mandinach, Honey, & Light, 2006), infrastructure investments (Gramlich, 1994), environmental and public hazard identification (Lawson, et al., 1999), disaster impact assessment (Lütfi Süzen & Doyuran, 2004), biodiversity conservancy (Maddock & du Plessis, 1999), emergency services (Lee, 2007) among others. Foreseeing the potential that data can have on our deeper understanding of urban systems, Batty (2008) described the emergence of a new urban science based on empirical data that acknowledges the complexities of cities and the myriad of interactions that happen inside them.

Nowadays, planning professionals increasingly use open data sources to do their work (Chakraborty, Wilson, Sarraf, & Arnab, 2015). Open data is often indispensable for public policy development and service delivery (Janssen, Charalabidis, & Zuiderwijk, 2012). These open data sources provide data that is free to use and disseminate, and mostly generated by government agencies in top-down processes (Janssen K. , 2011). The accessibility of open data can also foster a hacking or open source culture (Baak, 2015).

Notwithstanding the benefits of urban data, open data and data-driven decision making, challenges remain, including: unrepresentative and non-standardized datasets, varied software and server platforms to analyze it, and information gaps not addressed by available datasets (Chakraborty, Wilson, Sarraf, & Arnab, 2015). These can be particularly problematic for organizations with lower planning capacity and resources (Donovan, 2012). Furthermore, even when they exist, data are often not accessible for non-governmental agencies and researchers (Rambaldi, Kwaku Kyem, Keith McCall, & Weiner, 2006).

The quality of, and access to, data are especially limited in areas such as informal settlements (Zetter & De Souza, 2010). Such places can thus be excluded by formal planning processes and management efforts, with the lack of data used as an excuse (Werlin, 1999). In theory, then, open data resources can be key for understanding urbanization and integrating informality into urban management and planning processes (Chakraborty, Wilson, Sarraf, & Arnab, 2015).

2.2 Participation in smart cities

Technical developments are eroding the boundaries between planning and smart cities (Batty, et al., 2012). As data is continually updated in real time, planning could become more and more about continuous incremental changes rather than long term planning (Kitchin, 2014). As seen by many, particularly by industry vendors, smart cities are automated monitoring and control systems that allow cities to work more efficiently (Goodspeed, 2014).

According to the rhetoric, with the involvement of Information technologies (ITs) in more urban activities, sustainability problems can be ameliorated by improving efficiency and sometimes revolutionizing urban services, as is happening, for example, with smart grid energy management solutions (Carvalho, 2014).

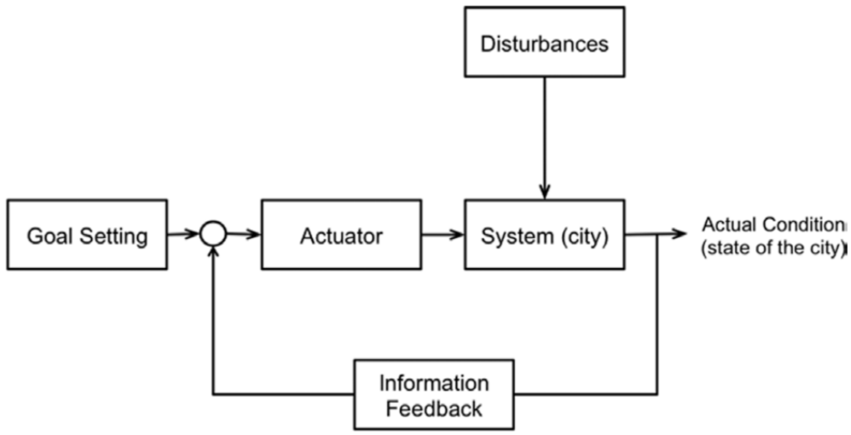

Figure 1 – Urban Cybernetics model (Savas, 1970).

Goodspeed (2014) equates the commonly accepted concept of a smart city with the urban cybernetics model (Figure 1) developed by Savas (1970). In this model, the city system is optimized by a combination of sensors, human or not, to gather data and create information feedback and urban models that can predict the most efficient decision to take in any situation. The model highlights the information feedback loop that an optimization and control model needs to be aware of the current state of the system and the consequences of its actions. Savas identified challenges to achieving a system like this, including disagreement about goal setting, government management problems and the inherent complexity of urban systems. Goodspeed characterizes related urban issues as “wicked problems”, with no clear definition, where value judgements play a large role and where setting goals and tracking progress are difficult to achieve (Goodspeed, 2014). Problems like equity, accessibility, mobility, and disaster preparedness and relief can be described as wicked problems where each stakeholder has a different appreciation and understanding of what the problem is, the current state of its reach, the goals to solve the problem, and how to measure the solution.

A smart city that functions following a centralized and automated computer-based governance system that operates all the urban subsystems is a concept mostly abandoned in the practical sense. Nonetheless, ambitious, large-scale smart city plans exist, including Singapore's iN2015 (intelligent nation) project, South Korea's business city, Guangzhou Knowledge City in China and Masdar City in the UAE (Hollands, 2014). Elements of automated monitoring and control are often adopted as "partial" smart city solutions to specific short term issues like traffic congestion and energy management (Goodspeed, 2014) (Carvalho, 2014). These approaches tend to reflect top-down creation processes, with great input from entrepreneurs, experts and engineers that set the systems in place, but little room for participation from urban denizens (Goodspeed, 2014). As Goodspeed argues, however, "no amount of innovation can avoid the necessity of making hard collective decisions" (Goodspeed, 2014). Carvalho also warns: "At best, a top-down 'smart technology' push is likely to be insufficient" (Carvalho, 2014). This can lead to an expert rule where the values of experts prevail over the values of residents (Rittel & Webber, M, 1973), leaving very little room for ordinary people to participate in the smart city (Hollands, 2014).

2.3 Community based alternatives to top-down smart city approaches

Figure 2 illustrates Arnstein's ladder of citizen of participation. In it, she illustrates a gradient of participation that goes from manipulating communities in non-participatory ways to full citizen control by an empowered citizenship. As mentioned earlier, most common understandings of the concepts of urban data and smart cities exclude citizens from the process of generating, analyzing, understanding, and making decisions with data. On the ladder, urban data is mostly used as a form of informing and consultation, while many understandings of smart cities do not open any spaces for direct intervention from citizens.

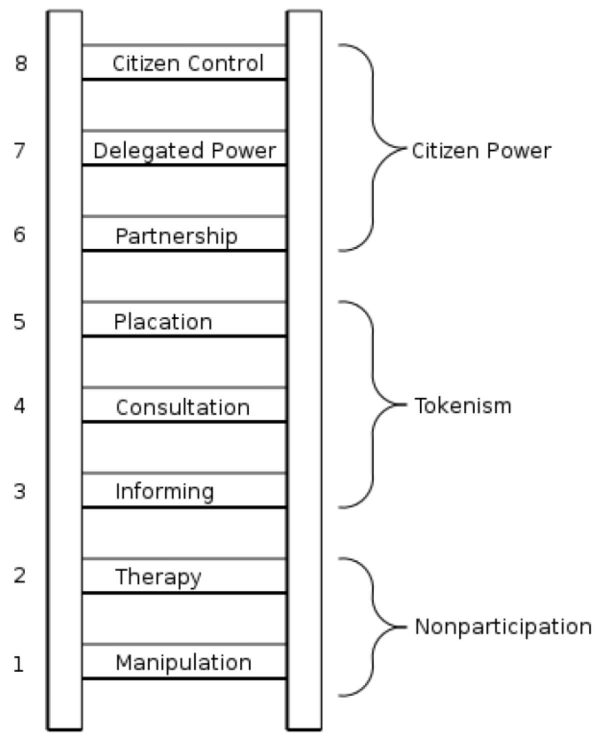

Figure 2 - Eight rungs on the ladder of citizen participation (Arnstein, 1969).

In their study, (Alawadhi & Scholl, 2013) found that the concept of smart cities is associated not only with data and automation in the minds of government officials, but also with transparent governments and people-centric approaches to technology. The focus on people-centric approaches hints at a governance level understanding of the need for public input and participation when modernizing the way cities tackle their problems.

Goodspeed (2014) proposes two approaches to achieve IT-enabled collaborative planning: local municipal innovation and the incorporation of IT into collaborative urban planning. This is an approach that has been increasingly adopted in recent years, with IT-based approaches that favor community engagement, participation and local innovation (Carvalho, 2014). Greenfield imagines digital urban technologies that allow for a 'spontaneous order from below' (Greenfield, 2012). Towards these ends, many local governments, aiming to create more people-centric digital innovations have established local innovation agencies that deal with the issue of coordinating the digital efforts of government agencies, companies, and citizens (Goodspeed, 2014).

2.4 Potential benefits of democratizing digital data collection

Community-based data collection can serve as a participatory counterbalance to top-down data sources. If community-created open data was collected in a variety of subjects and easily available, the planning process could have more evidence-based data sources to inform its decisions. Community-generated data could complement, confirm or challenge the data generated in a top-down manner. For this, digital data collection needs to be democratized: that is, making the technologies to collect urban data, and the techniques to use them, as widely available as possible.

Some potential benefits to democratizing data collection in our societies include:

- To make gathering of tailored datasets more commonplace to help solve urban wicked problems;
- To complement the top-down smart city movement and big data flows with bottom-up place-specific data created by the same communities being studied rather than by external agents;
- To make it simpler to fill in data gaps created by informal settlements and systems;
- To enable faster, community-based, data-driven disaster response.
- To help smaller community-based organizations undertake data-driven efforts to understand and improve their communities; and,
- To make it simpler for local authorities to update outdated, or infrequently updated, datasets (e.g., census data).

2.5 Strategic niche management for embedding new urban technologies

Strategic niche management is a method to facilitate the introduction of new technologies originally developed for sustainable technologies (Caniëls & Romijn, 2008). It starts with small experimental interventions that help to understand and develop a technology in specific niches with the final goal of embedding the new technology into the sociotechnical landscape (Figure 3) for general use (Schot & Geels, 2008). Carvalho (2014) summarizes the work of Truffer (2002), Kemp (1998) and Schot (2008) on the sociotechnical theory approaches of strategic niche managing for successfully developing and ultimately embedding new technology into societies via two key activities:

- Learning – Testing and fine-tuning technologies, their variants and conditions of success in real-life environments.
- Societal embedding – The progressive interaction between new technologies/solutions and the social, cultural, political and governance dimensions that structure their use, including three interlinked processes:
    - Network building - The creation of constituencies and coalitions of public and private supporters of the technology (potential producers, users, regulators) and resource pooling (for example, money, expertise);
    - Infrastructure matching - The adjustment of the new technologies to an existing sociotechnical environment of regulations, standards, business models and physical artefacts; and
    - Expectation building - The development of favorable expectations and visions about the advantages of the new technologies for society, (international) attention, and legitimacy for continuing experimentation.

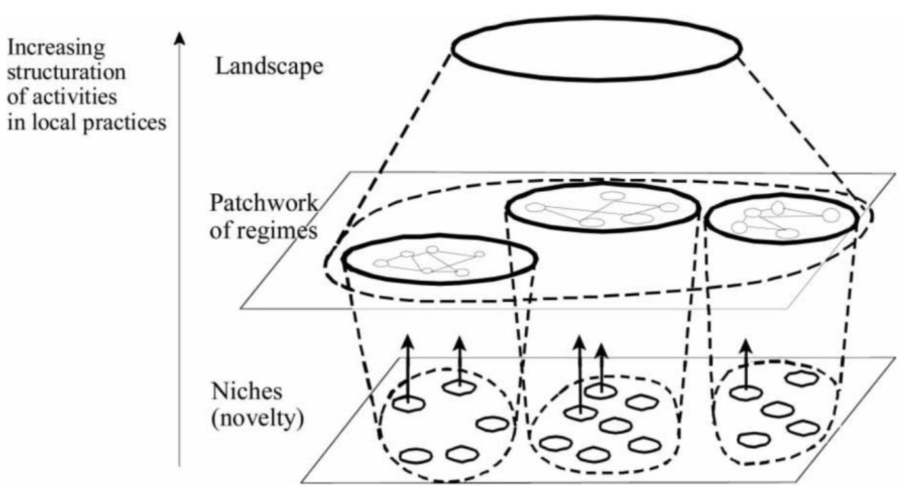

Figure 3 – Different levels of the sociotechnical landscape Schot (2008) based on (Geels, 2002).

2.6 Digital community mapping as a niche for democratizing data collection

Community mapping is the act of creating maps with the members of a community. Involving people in the mapping process can be a subversive bottom-up activity (Pinder, 1996). The act of community mapping can focus heavily on the activity itself, not on the map (Harley, 1989), which suggests the importance of context and process, rather than outcomes (Perkins & Thomson). The involvement of both the community and the organizers in the creation of the map can be a form of Participatory Action Research (Amsden & VanWynsberghe, 2005). Digital community mapping is a form of "volunteered geographical information" (VGI) (Goodchild, 2011). VGI is a primary source for open urban data that can be used to supplement, and fill gaps left by, more formal data sources like census information or satellite imagery (Chakraborty, Wilson, Sarraf, & Arnab, 2015). It is a manner to crowdsource information about a place (Crooks, Pfoser, Jenkins, & Croitoru, 2015).

Digital community mapping is a way for planners to get more in touch with local residents' understanding of a place and for local people to reflect about their reality and create a meaning of the place themselves. For many community mapping projects, very little resources are available, and pen and paper are the tools utilized, whilst more wealthy places can afford time and money for developing custom technological solutions. There have been some large-scale purpose-specific digital community mapping projects (Open Street Maps, etc.) and relevant customizable tools exist, some community-based (Open Data Kit, Ushahidi, OpenPlans) and some professional-oriented (Survey123, Fulcrum).

2.7 Flocktracker

Among the technological innovations allowing for new venues of participation, Smartphones are remaking both the creation and consumption of information by allowing users to engage in mapping, photography, audio and video capture, capture of GPS sensor data, etc. (Boone, 2015). Flocktracker is a smartphone-based data collection platform that aims to eliminate the technical barriers often faced by researchers, urban planners, and communities when gathering data in the field. It tries to automate as much as possible the process of creating a new data collection project while maintaining the users' control of the content of the data collection. It also aims to be as simple to use as possible, so that people without technical backgrounds can set up and deploy their digital field data collection efforts without previous training.

Consistent with the concept of strategic niche management, Flocktracker has been used in numerous research projects around the developed and developing world, allowing users with different levels of technical expertise and purchasing power to collect the data they need for their goals. These projects include collecting data on semi-formal public transport data (Ching, Zegras, Kennedy, & Mamun, 2012) (Zegras, et al., 2015), perception mapping of transit users (Butts, 2014) (Zegras, Butts, Cadena, & Palencia, 2015), environmental planning of public parks (Hwang & Roscoe, 2017) and others (Flocktracker, 2018). Building from these encouraging examples and the potential benefits of widespread usage of digital data collection tools, this thesis aims to examine the possibility for societally embedding Floctracker. Although still based on niche experiments with key partners, the research focus moves beyond technology matching to attempting to understand Flocktracker's broader potential: in assisting societal partnerships, helping to develop strategies in the urban planning arena, and in democratizing data collection such that it can be a common part of the societal landscape. In the case of this study, the societal landscape is defined as communities in the developing world.

2.8 Summary

Realizing the true potential of the “smart city” movement requires broad engagement and empowerment of the communities themselves. Digital community mapping is one activity that combines both community participation and the creation of data. These two activities match well the objectives of Flocktracker, and the digital community mapping niche already shows, albeit at a small scale, the benefits that democratizing data collection technologies can have on organizations related to city planning and their communities. This thesis examines attempts at digital community mapping in developing country contexts to identify the potential institutional and community interactions that could help foster the democratization of data collection and embed it in the sociotechnical landscape of a place.

3 A field data collection tool: Flocktracker

This chapter describes the design processes and drivers of Flocktracker, including a description of its user experience and components and a high-level technical overview of its implementation details. It also presents one proposed model of what democratization of the tool might look like for users and society more generally.

3.1 User experience design drivers and process

Flocktracker is a flexible data collection platform initially created to be easy and cheap to deploy in a variety of contexts, leveraging smartphone technology. Over time, Flocktracker evolved to become a content-agnostic data collection tool, designed to provide an easy-to-deploy, easy-to-manage, ad-hoc, low-cost, cloud-based field data gathering method. With Flocktracker, high-resolution field data collection projects can be rapidly set up and deployed. The technology enables a range of users, from citizens to academics to businesses, to collect highly heterogeneous field data, store and manage it in the cloud, monitor data-collection in real-time, and, ultimately, choose to share project results among a wide range of interested parties. The tool is envisioned to enable innovative research approaches both in the quantitative and the qualitative realms in a data collection design as described in section 3.5.

Flocktracker's design process has been iterative, starting in 2011, using field experiences and needs as the main inputs for the definition of design and implementation features. It was first deployed to create the First Bus Map of Dhaka, in a collaboration with an NGO in Dhaka, Bangladesh (Ching, Zegras, Kennedy, & Mamun, 2012) (Zegras, et al., 2015). A second implementation in Mexico City, Mexico (Zegras, Butts, Cadena, & Palencia, 2015), which aimed to map bus routes and carry out on-board surveys of bus users, entailed a large amount of software development time to adapt to the new context and new data needs. These techniques were later streamlined in St. Louis, United States (Butts, 2014). These early experiences made it clear that an important step forward in allowing anyone to embark on their own data collection processes was to provide a flexible software framework to easily develop and deploy projects with as little technical knowledge as possible. Since these initial implementations, Flocktracker has been used in nearly 30 field data gathering projects on a range of topics and implemented by a range of users.

Flocktracker's user experience design is based on User Centered Design (UCD, also Human Centered Design, Figure 4) and the Android Material Design paradigm (Google, 2018). The following list contains some of the main drivers of the design.

- The existence of two main types of users.
    - Tech-savvy project entrepreneurs that are more aware of the inner workings of the tool and know how to set it up and spread it to their peers. They have more in-depth knowledge about basic data tools and productivity software. They are the main drivers of data collection projects and will be primary users of the data generated, thus they have more motivation to learn about how to setup the software. Although the intent of the experience design for them is to be able to setup projects by themselves without prior knowledge, a user manual was developed to help with the process.

  - A more general audience that does not need to understand deeply how the tool works but has basic knowledge about how to use popular cellphone apps and web services and basic productivity software knowledge. They are not necessarily motivated to learn about, nor have prior knowledge of, how to setup Flocktracker, but they are interested in collaborating with data collection projects, either by their prior relationship with the project creator or because of their interest in the data collected.
- The tool has been developed with flexibility and easy learning curves in mind.
- It should work as closely as possible as existing software on mobile devices. Although the complexity of the functions in the program is closer to what desktop apps normally offer rather than the often-hyper-focused functionalities of mobile apps, the design aims to be as simple to use as possible for users familiar with mobile apps. This was achieved by having screens, with single focused functionalities, back button navigation, large self-descriptive icons and a pleasing design.
- The software should have high resilience to software crashes. Even if the app fails and must restart, it should resume exactly where the user left her work. The app should "feel" sturdy and reliable.
- The app should be resistant to internet connectivity shortages. The app should keep working seamlessly even in places with no internet connectivity and should upload any data collected to the server as soon as it has access to the internet.

- The selection and development of new features should come from understanding current challenges to using the tool as expressed by the users themselves or by analyzing usage data of the platform.
- Qualitatively, its intended complexity of use sits between that of general audience social media and basic office tool platforms like Facebook, Twitter and Microsoft Word; and professional geospatial software that requires greater expertise like ArcGIS and QGIS.
- It should allow for a large variation in terms of the number of people collaborating on a data collection project, from small teams composed of closely related individuals to large-scale crowdsourced data collection projects.

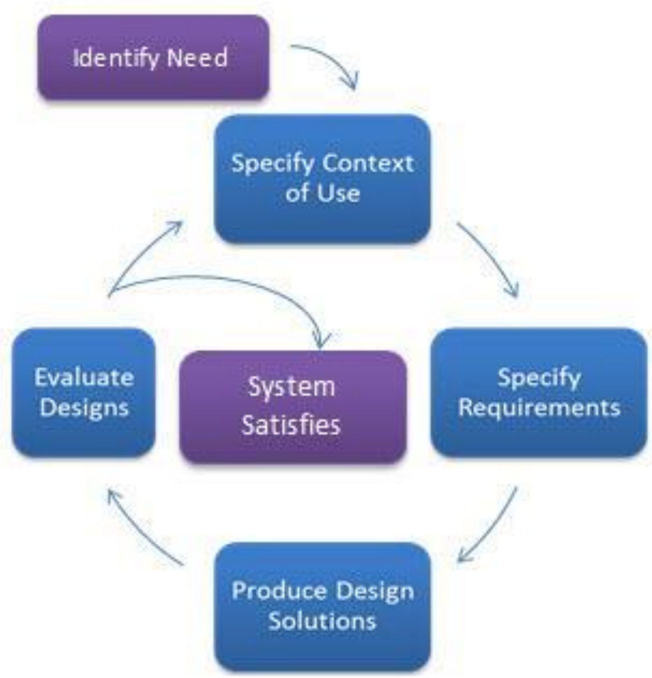

Figure 4 - User Centered design process (U.S. Department of Health & Human Services).

3.3 The in-field data collection workflow

This section describes an abstraction of the steps of an urban data generation process, from ideation to results that are meaningful for learning and taking action. These processes can range from doing pen and paper surveys outside a subway station about quality of service to setting up sensors around a town to gather data about air quality. The process described was the inspiration for the overall design of the Flocktracker platform and the community experiment design.

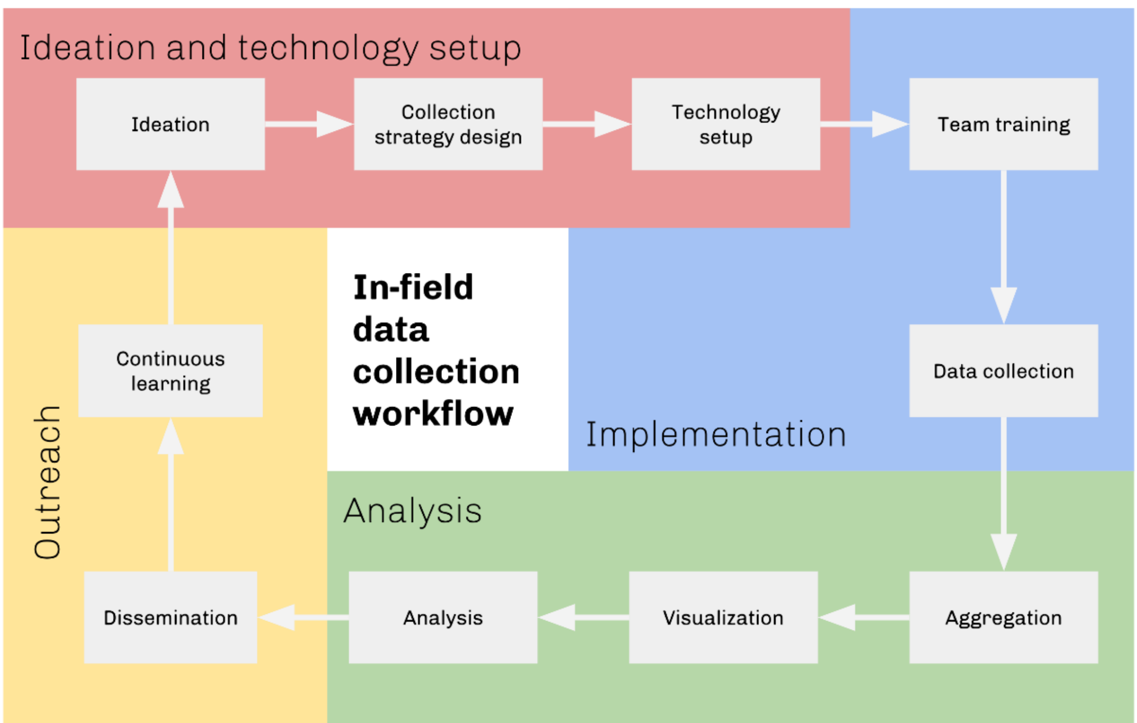

Figure 5 – Stages of an urban data collection project and its steps.

As the diagram in Figure 5 suggests, the envisioned process is a self-feeding cycle in which people that create data collection projects and collect the data repeat the process for different data collection needs. This iteration can help to make the process more streamlined as the project designer develops more experience implementing these projects and learning from other related projects. Its similarity to the Urban Cybernetics data feedback loop model (Savas, 1970) is intentional. The process consists of four stages – ideation, implementation, analysis and feedback – as detailed below.

Some steps inside each stage might overlap or even combine some projects depending on available resources and strategies to finalize each stage. This is very common in the Ideation and Technology Setup stage, where the possibilities of what to collect are very much related to the technologies and strategies chosen. For example, people that only know how to use pen and paper as their data collection technology will not be able to get accurate GPS data for locations but might decide to print maps and mark places on them to get a sense of the location of their data collection.

3.3.1 Ideation and technology setup

This stage of the process includes going from an idea to a technical solution that will allow the data collection to happen. This can be an iterative process itself, since the potential data that can be collected is limited by the capabilities of the design team and the potential participants and the capabilities of the technology accessible to the team. A team might want to use a digital map to draw all the routes that people use when walking around a place, but if they do not know how to set up the map, or if they think that their community will not know how to use it, they might change to using pen and paper to draw the routes on printed maps. This, in turn, will limit the potential to collect data with the community to in-person outreaching or mailed-in maps asking people to send their answers. This change in media and technology will then be taken into account in considering the limitations of the new pen and paper system, where pictures are not easily attached, and future digitalization of the data collected for analysis and publication can be time-consuming. This iteration will continue until a good compromise between the quality, scope and contents of the data to be collected is achieved based on the technologies available to the participants in the design of the data collection process.

3.3.1.1 Ideation

Based on specific needs that come from previous projects, literature and the planner's understanding of data needed to ease understanding of a problem in the field, a project is ideated. The ideation and definition of what data to collect can come from the need to fill-in institutional data release gaps, to update already old information, and/or to explore new subjects in the field. This could be, for example, collecting all the informal bus routes in Dhaka, Bangladesh or making a survey of all the public sanitation facilities in Solo, Indonesia. The ideated dataset might include data that needs to be collected with sensors or other measuring devices, like location data, acceleration data, and/or exact time and/or place of data collection. Some projects might just need survey information, while others might require GPS location data to be useful and others might need data to be collected at a specific times and locations, like when collecting all the bus route data of a city.

3.3.1.2 Collection strategy design

Once the project is defined, a strategy on how to use the available resources to collect the data must be developed. This can include using traditional technologies like pen and paper, making alliances with organizations like universities or companies with the know-how to collect this information, using tools already available in the market, or creating tools purposely developed for the project. In this step, the limitations of the selected technologies start becoming evident.

3.3.1.3 Technology setup

With a strategy for how to solve the data collection technological needs, the technology is either developed or set-up for deployment. This includes prototyping and testing the tools to make sure they will work when deployed in the field. Technologies can vary from pen and paper and surveyors in the field, off the shelf sensors to deploy around the city, or smartphone apps that allow for data to be collected by a team in the field. As mentioned before, the selection of the technology to be used often influences the ideation of the data collection project itself, which creates an iterative design loop.

3.3.2 Implementation

During this stage, the project designers meet with the people that will collect the information, if they are not doing it themselves, and get the information needed about the field. As digital data collection on the internet allows for near real-time monitoring of the process, changes and corrections can be done while the data is still being collected. This is the step where direct public participation has most often been seen in the case of Flocktracker and in community digital data collection projects. Participation from the public is also easier in this stage since it is the one that requires the least specialized knowledge.

3.3.2.1 Team training

Once the tools are ready, or sometimes during the prototyping and development phase, the training of the data collection team takes place. This includes making sure that all the information to be collected is understood by data collectors, to minimize misunderstanding and systematic errors, and the last stage of testing of the technology and the data collection project itself. In the case of sensor data collection, this process includes debriefing with participants about how to deploy the technology and keep it working. This is the last opportunity for the project's designers to identify problems with the design of the project, the implementation of the technology, or the nature of the data collected itself.

3.3.2.2 Data collection

This is the step, in which the team collects data in the field. The duration of this stage varies with the amount of data to be collected, the comprehensiveness of the study at hand (from exploratory data collection to comprehensive surveys of a place), the size of the team, the need or lack thereof for monitoring temporal changes in variables collected, and the complexity of the data to be collected. The possibility for live monitoring during this step can be of great value, for: making the data collection more efficient and comprehensive; understanding the evolution of data collection over time (e.g., to observe variations in data collection by different surveyors due to comfort levels with the tools and/or interest in participating); and, monitoring and auditing surveyor activity (in near real-time).

3.3.3 Analysis

In this stage the data is compiled, aggregated, visualized and analyzed. Depending on the success of each of the steps in this stage, there might be a need to backtrack to create different visualizations or manipulations of the original information to enable more meaningful analysis. This stage also involves an evaluation of the success of the data collected in fulfilling the initial project objectives (as established in the design stage).

3.3.3.1 Aggregation

This step entails aggregating all the data collected from the relevant sources (like sensors and people) and sometimes linking these data with other data sources. This step can be time-consuming if, for example, traditional pen and paper surveys are used. Standardization of the formats for collecting the data is key to a streamlined aggregation process. Many digital tools to collect data, like online forms (Google Forms, Survey Monkey), do this aggregation on the fly, making this step automatic.

3.3.3.2 Visualization

Data visualization can be done with maps and charts to help provide further insights on the data collection process, allowing the data collection team to draw initial conclusions. Several relevant tools exist to ease this task, both for free, like QGis or D3, and as paid services, like Microsoft Excel, Mapbox, ArcGIS and Carto.

3.3.3.3 Analysis

The analysis of the information can be done based on the visualizations produced, geoprocessing, statistical methods (e.g., econometrics, machine learning) and other tools that will vary depending on the questions to be answered, the savviness of the analyst, and the complexity and consistency of the data. As with visualization tools, analysis tools that facilitate this step of the work are readily available and accessible, from open source platforms like R and Phyton to paid ones like SPSS and Excel.

3.3.4 Feedback

This stage involves using the outcomes of the analysis to inform the findings and conclusions, upon which relevant actions can be taken. Feedback comes from participants and other people that have access to the results, like colleagues, stakeholders and researchers. This feedback can lead to subsequent improvements in data collection processes.

3.3.4.1 Dissemination

The results (from the raw data to the final analysis) can then be disseminated through various avenues (e.g., publications and presentations) and to various audiences (internal and external). Based on the disclosure of the findings to relevant partners, actions can be taken, and further research can be strategized.

3.3.4.2 Continuous learning

Based on feedback from relevant audiences and the perceived shortcomings of the project, new, innovative ways of undertaking related data collection projects can be envisioned. This feedback can come from various sources, including comments in blogs and social media posts about the findings, and in various forms, such as technical and/or anecdotal suggestions. The goals are to improve the process and obtain better results.

3.4 Flocktracker: High-level technical overview

The Flocktracker platform is a cloud service with a server storing and relaying data to an Android app that acts as the data collector and visualizer. The server-client communications are based on the Representational State Transfer (REST, RESTful) web service architecture. When communicating with the server, the app uses Universal Resource Location (URL) to address each resource like users, projects and their components. Each resource can be accessed using different verbs (GET, PUT, POST, DELETE), each used to alter or access the data in a different way: GET to retrieve data, PUT to modify it, POST to add a new resource and DELETE to remove it from the server.

Communications are protected by three layers for encryption, authentication, and authorization. Encryption is achieved by the Transport Layer Security (HTTP over TLS) that uses the symmetric public key cryptographic algorithm to provide secure communications. Authentication is handled by the Basic Access Authentication method that allows for the server to verify the authenticity of the client's identity by attaching username and password data to every call to the server. Lastly, authorization to different projects and datasets is a user customizable feature of each project. Project creators can give or remove write and read permissions to other users. Only when the three layers are acceded successfully, the server processes the request of the client.

The app can display the Project Builder that allows the user to manipulate the data structures needed to define a project; this data structure is then sent to the server which, in turn, sets up all the databases needed to start collecting data. The builder is made to comply with the FlockSON[1] survey markup language. This compliance aims to remove potentials for user error in the fulfillment of this markup language's requirements. User management involves graphical features that allow for different levels of authorization to be given to different users.

Data collection in offline conditions uses a data cache queue service that stores locally all the surveys, counters and tracker traces. Another service attaches GPS location whenever a new upload is generated. Once the app detects that the user is connected to the internet, it starts emptying the cache by sending data to the server. Only after the server has confirmed that the data is properly stored are the data on the cache deleted.

Data visualization occurs in a customizable map that allows the user to filter the data according to the different survey questions, counter elements, and/or the specific user that uploaded the data. This interface also allows for the download of all the data in CSV format for further post processing and gives the user the ability to generate hyperlinks (URLs) to download and share all the data on other devices.

[1] FlockSON (Flock Survey Object Notation) is a markup language developed for Flocktracker based on the JSON language. It is designed to handle the definition and transference of surveys with complex data structures and data collection flows.

3.5 Components

The user-facing side has four main components (Figure 6) that are intended to be embedded into the data collection process to make it easier and seamless.

- Project Builder - A customization tool that allows the user to design, build, edit and test different data structures before deploying them and starting to collect data. The user can design his or her own projects by adding the needed data structures and questions to the project. The project can then be uploaded to the server to share it with others and allow them to start collecting data.
- Data Collection - This component allows for collaborative data collection by several projects concurrently, including GPS movement tracing, surveying, and counting interfaces. Users can setup different levels of permissions for the participants in their projects, allowing them to control who can upload, download and manage data.
- Data Visualization - This allows the user to visualize maps and charts of the data as soon as they are collected as well as the location of the data collection team. Not only can reports of the data collection be generated, but potential biases in the data collection process can be identified in real-time and the data collection team can be more efficiently managed on-the-ground.
- Exploration – This component intends to facilitate inter-project collaboration by allowing the exploration and finding of datasets and data collection projects related to any user's location and interests. This function is not yet implemented and thus was not

part of the interactions people had with the Flocktracker platform during the research of this thesis.

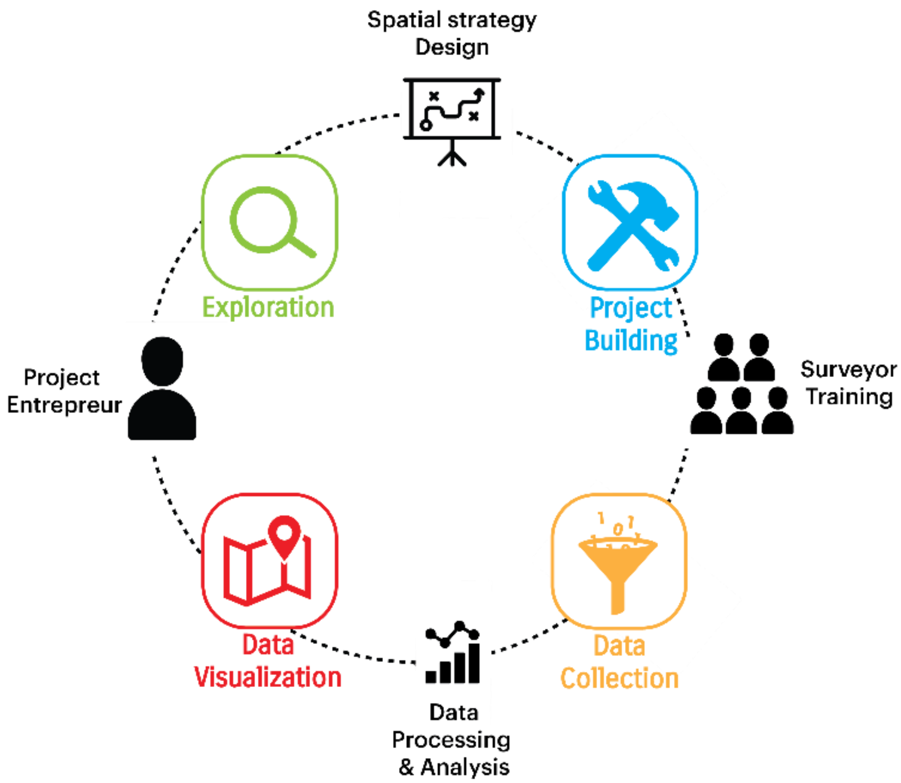

Figure 6 - Flocktracker components embedded in the data collection workflow. Diagram by Chaewon Ahn based on Arturo Cadena's initial work.

3.6 Data Structures

Each Flocktracker project can have any combination of these three components:

- Survey - The survey component allows for as many questions as needed, of many types. The survey format is based on FlockSON. Surveys can include several different kinds of questions, with internal logic (e.g., for conditional branching), and can contain pictures.
- Counters - This basic tool allows for relevant elements (e.g., traffic and vehicle occupancy counts) to be counted in the field.

- Tracker – The tracker traces the movement of the device by polling the GPS sensor location every 15 seconds. It can also allow for survey questions, also defined with the FlockSON survey markup language, to be asked before and after starting the tracker. This has a range of potential uses such as mapping informal transit, preferred routes, and basic activity pattern data collection.

All the data collected with any of these structures is tagged with username, timestamp and GPS location.

3.7 Limitations

The software has at least four limitations that could limit the potential participation of community members.

1. The software is not well integrated with social media. It does not allow for sharing the content generated in forms other than URLs or files. This prevents leveraging social networking platforms to outreach directly to the Android app.
2. The app only works on the Android platform which prevents people with iOS from participating. Although Android is the prevalent smartphone platform in the world, iOS is widespread enough to justify support. This shortcoming may have greater implications in the United States and Europe, where iOS has a larger market share. Presumably, in the developing world, where this thesis research takes place, Android is dominant thus reducing the consequences of this limitation.

3. Data visualization currently consists of maps that the user can filter based on the data collected. Users have, however, identified the need for charts to visualize data; this may be because people are more familiar with visualizing data in a chart form.
4. The Exploration Tool is not enabled in the platform. Users are currently not able to see projects that they were not part of nor join projects that they were not invited to explicitly.

3.8 Democratizing data collection – A model within Flocktracker

The potential for spreading the use of the Flocktracker platform in the sociotechnical landscape of a community relies on the ability of Flocktracker to allow users to join each other's projects to foster the creation of data collection communities. This is how 'spontaneous order from below' (Greenfield, 2012) would look like in the Flocktracker platform. To test this potential, I have conceptualized a four-stage model which entails: project seeding, initial and secondary spread of the tool's usage, and, finally, establishment of the tool within the community.

3.8.1 Model assumptions

The Flocktracker "democratization" model rests on some assumptions about the available technologies and the community itself, namely:

- The usage of the data collection technology can be taught in one short introductory session;
- Widespread access exists to the platform the technology relies on, or, in the case of Flocktracker, access to Android devices and cellular internet connections;

- People in the community can learn about the usage of the tool;
- In-place community organizers have capacity to organize and setup a data collection process;
- An interesting enough subject exists for which the community can be engaged and motivated to collect data; and,
- This community is not generally averse to data and collaborations with community organizers.

3.8.2 Democratization stages

Below is a description of each of the four democratization stages envisioned and a brief discussion of some metrics that could help to characterize and study the democratization process moving forward.

The seeding stage is envisioned in this thesis as an initial phase within the niche of digital community mapping, but different seeds could be designed in the future to match different niches (e.g. citizen science, political advocacy, nature conservancy, climate change resiliency, disaster relief).

In practice, the observation of these different stages can be difficult since they may happen in tandem once the initial seed project's outreach campaign stars. For example, many people might setup the platform at, or very close to, the same time and start collaborating on their own projects creating an in-place network from the beginning.

3.8.2.1 Seeding stage

First, community organizers, ideally with participation of their communities, must identify a subject deemed of interest for data collection. Following a guided process to allow for community organizers to familiarize themselves with the tool, a data collection process is setup on the Flocktracker platform. This stage requires the most resources since personal interaction is key to success. The project should be as simple as possible, with just a few questions and setup steps to facilitate its further explanation and eventual spread.

The intended geographic scope of the project should ideally coincide with that of the normal activities of the community members; this will allow for people to participate without deviating too much from their usual paths. This poses the risk that data collection will be concentrated in the places people know and are interested about, which would reduce the quality of the data collected potentially making it irrelevant, but it might also increase the chances of additional participation and enhance the quality of the data as people would be reporting about places they know (Millar, Hazell, & Melles, 2018). A balance between the competing needs of data quality, geographic spread, and ease of access to encourage more participation should be considered.

An initial timeframe for the community data collection should also be set to allow for the data collection process to be better focused. After the established time has passed, a second event can be held where people share their learnings about and new understandings of the subject of the data collection, the data collection process itself, and urban data in general.

The project is then shared with a broader audience in an initial community meeting. This event can be advertised via the internet and social media, leaflets, posters, and in-person communication. In this initial stage, the organizers should closely accompany community members to make sure that they understand the value of generating and keeping track of urban data and the technical process of how to collect data. People especially interested in the process are also trained on how to design and setup data collection processes themselves. In Figure 7, the Flocktracker logo represents the "seed" project that is setup with a community.

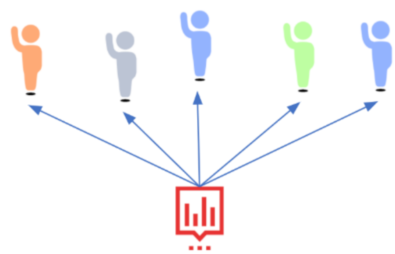

Figure 7 - Seeding stage – A project is created and shared with the community.

For this stage to be completed successfully, members of the community must be taught how to join a project and to collect data themselves. Although the design of Flocktracker is intended to be easy to use, it would be ideal to have at least some interested members of the community trained on setting up projects themselves.

After the training, a period to invite people to collect data for the project on their own starts. People can be incentivized to collect more data. These might include prizes to the people that collect the most data or partnerships with local authorities to act on the data collected.

Some metrics to observe different aspects of this stage include: the ratio between people that attended and total members of the community; the ratio of people that understood how to setup the platform on their phones and people that attended the event; the average number of surveys per person during the initial community event and in the data collection time window after the event; and the geographic spread and density of the data collected.

3.8.2.2 Primary usage spread stage

In this stage (Figure 8), some of the participants of the seed project might want to setup projects on their own to collect data about other subjects of interest. Every new project would operate independently. Relevant metrics to observe this stage include: the number of projects created in a place by people that participated in the seed project; the number of people that joined the seed project; the time that passed between the first community event to share the seed project and the creation of the first and subsequent projects; and, the ratio that divides the number of data points that each user collected in the seed project by the average number of projects created by the same user. The geographic spread and density of the set of projects created can also be measured to identify places where data communities are more interested and active.

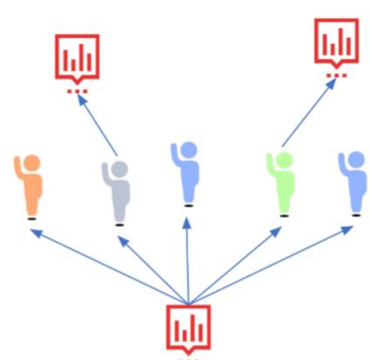

Figure 8 – Primary usage spread stage – People create their own projects.

It could also be illuminating to identify the “subject spread” of the subsequent projects created, or the similarity of the new projects’ subjects to that of the seed project’s subject. This could be qualitatively assessed. For example, if seed project A is about mapping local businesses and subsequent project B is about local restaurant menus, and subsequent project C is about mapping defective street lights, B would be closer to A than C. Quantitative measures could also be used here, using, for example, a metric of semantic similarity via data analytic methodologies. A larger spread and variety of subjects might indicate a more data savvy community. A small spread and variety might indicate that the initial seed project's subject influences the subsequent use of the platform.

3.8.2.3 Secondary usage spread stage

In this stage (Figure 9), people that did not participate in the original project adopt Flocktracker and create their own projects. In this stage, members of the community might or might not be aware of the seed project and its content. They just use the platform to fulfil their data collection needs.

Various metrics can be used to assess this stage. If, in terms of subject matter, a range of projects were generated in the previous stage, then we could measure the relative success of each of the different subjects in attracting new people to join a given project, collect data, and create new projects. With many projects created, the ratio dividing the total data points collected by the number of people participating in each project can be measured for comparison among projects, subjects, and geography.

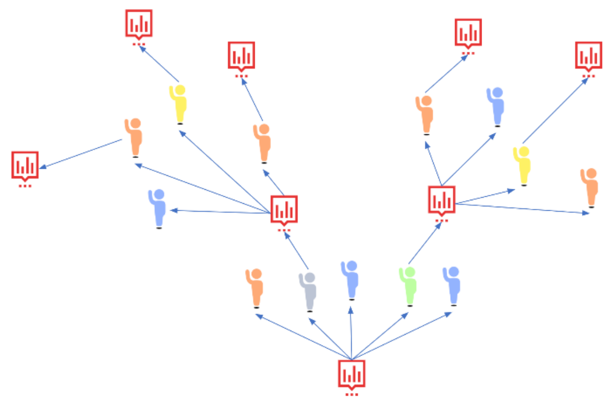

Figure 9 - Secondary usage spread stage – New people join and create their own project.

3.8.2.4 Community establishment stage

In this stage (Figure 10), the tool becomes established in the sociotechnical landscape of a community. Different users would be part of several projects and data collection exercises and they would be able to generate new data collection projects as they see fit.

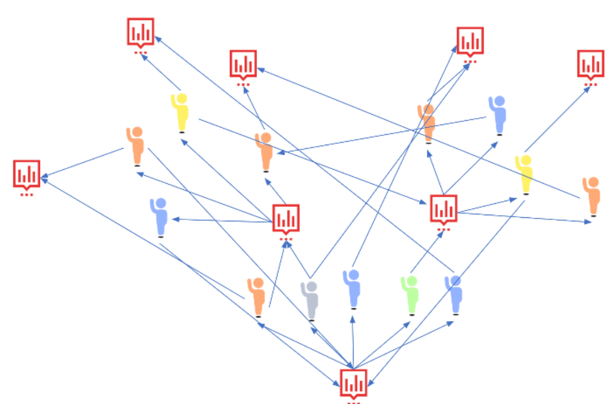

Figure 10 - Community establishment stage – A network of active projects is established with the community.

Aside from the metrics outlined in the previous stages that compare different projects, the "health" of the network can be monitored by keeping track: of the number of people accessing the different datasets associated with the different projects; the number of uploads per project; the number of people joining the different projects; the average number of people in each project; the average number of projects each person is part of; and, the average number of surveys in each project. Network analysis would allow for the identification of sub-communities inside the network based on participation in projects with similar subjects or on mutual participation in the same projects.

With many projects, the data density of the network could be measured to allow for identification of potential biases when collecting data in the community. Other metrics include the average active data collection life of a project, average number of downloads of each of the project's respective dataset, and which subjects have the most dataset downloads.

4 Research Setting and Methods

4.1 Introduction

This chapter describes the methods designed to answer the research questions outlined in the introduction, regarding: (1) the utility of Flocktracker, the tool, as well as the "democratization" model explained in the previous chapter; (2) the effects of participating in the data collection process on individuals' perceptions regarding data and participation; and (3) whether an initial community data collection experience generates new data collection activities (i.e., does a "seed" project lead to more widespread usage of the data collection tool). To answer these questions, I co-designed and implemented four experiments in partnership with two community advocacy organizations, in two different places in the Global South: Surakarta (Solo), Central Java, Indonesia and Tlalnepantla, Greater Mexico City, Mexico. The experiments also shed some light on the role that different types of local partnerships and different outreach techniques can play in spreading the use of digital community mapping tools in a community, including the number of people reached, and their participation.

In the following sections, I describe the empirical contexts where the experiments were carried out, and the research methods used to answer the primary research questions, which included pre/post-surveys for assessing impacts of the workshops on the participants and instrumentation of the Flocktracker platform in order to observe the stages by which participants adopted the technology. I also describe my own assessments of the experiments based on participating in and observing each organization's process of implementing Flocktracker, interviewing the participants, and evaluating the outcomes of the four projects in relation to each organizations' needs. Finally, I consider how my direct involvement in the experiments influenced the process, shaped my opportunities for observing the process and outcomes, and informed my conclusions related to Flocktracker as a tool for fulfilling the communities' data collection needs and the organization of events like this as a way to spread the usage of digital data collection tools.

4.2 Empirical Settings

In this section I describe the experimental settings for this research, focusing on the similarities and differences of the cities where, and the organizations with whom, the experiments were developed.

4.2.1 Locations

Mexico and Indonesia are countries that experienced demographic booms in the second half of the 20th Century; for Mexico this trend peaked in the 1970's (Zavala de Cosio, 2014) with population growth rates of 3% annually (Cosio Zavala, 1995), and for Indonesia the peak was 2.7% in the late 1970's (United Nations, 2018). They both have experienced a trend of rural to urban migration, which strains city resources and has created large populations living in informality in both countries. Overall, Mexico is more urbanized: as of 2017, Mexico has a rate 80% of people living in urban areas, while Indonesia is at 55% (The World Bank, 2018).

Surakarta (also known as Solo) and Tlalnepantla (officially Tlalnepantla de Baz) are both located within the most densely populated regions of their respective countries: Tlalnepantla in the Megalopolis of Central Mexico (Corona Regional del Centro de México in Spanish)

(Delgado, Larralde, & Anzaldo, 1999), with 27 million inhabitants; and, Surakarta on the island of Java, the most populated island in Indonesia and the world, with over 145 million inhabitants (Badan Pusat Statistik (Statistics Indonesia), 2018). Informal settlements are common in both Tlalnepantla and Surakarta. Figure 11 depicts both places. Many of Tlalnepantla's neighborhoods are in the difficult terrain shown; these places are often ridden with crime and only served by private taxis and informal buses. In the Surakarta case, although the city is urbanizing fast, formally and informally, empty lots are still being used to grow rice.

Figure 11 - Left: Tlalnepantla's informal housing in Sierra de Guadalupe. Right: Rice paddy in Surakarta with a new building development on the background.

The Indonesian and Mexican governments differ in their structure and power allocation. Mexico is a federal state with local governments gaining increasing autonomy. Indonesia is a more centralized state which, while undergoing some devolution to local governments, still maintains a strong lead from national level politics and government. In terms of structure, Mexico has three levels of government: a) federal (gobierno federal), b) state (estado) and c) municipal (municipio). For Indonesia, legally, five levels of government exist: a) national, b) provincial (provinsi), c) regency (kabupaten) or city (kota), d) district (kecamatan) or sub-district (distrik) and e) administrative villages (desa) or urban communities (kelurahan). Surakarta is in the province of Central Java (Jawa Tengah), while Tlalnepantla is in the State of Mexico (Estado de México). Surakarta is composed of five districts (kecamatan) subdivided into 51 urban communities (kelurahan) in total, while the whole city of Tlalnepantla falls under one administrative division that is inhabited by more than half a million people.

A peculiarity of Tlalnepantla is that its territory is divided into two non-contiguous areas, the largest and more developed area that contains the city hall and most government administrative offices is located to the west and a smaller, less developed exclave is located to the east. They are separated by a mountain range, the Sierra de Guadalupe (Figure 12). This means that to travel between the two areas of Tlalnepantla, inhabitants normally travel around the mountains, outside the State of Mexico and into Mexico City before coming back to Tlalnepantla. The mountainous areas of Tlalnepantla present a major challenge to the provision of adequate transportation and other services. For example, a public transportation trip to the city hall from the informal settlements in the east area of Tlalnepantla on public transportation can take more than one hour and a half. Surakarta has a mostly flat territory that is crisscrossed by rivers and a network of canals used by the still active local agriculture. Its territory is consolidated and relatively easier to traverse with commutes within the city that rarely pass the one-hour mark since it is only 10 kilometers at its longest (Figure 12).

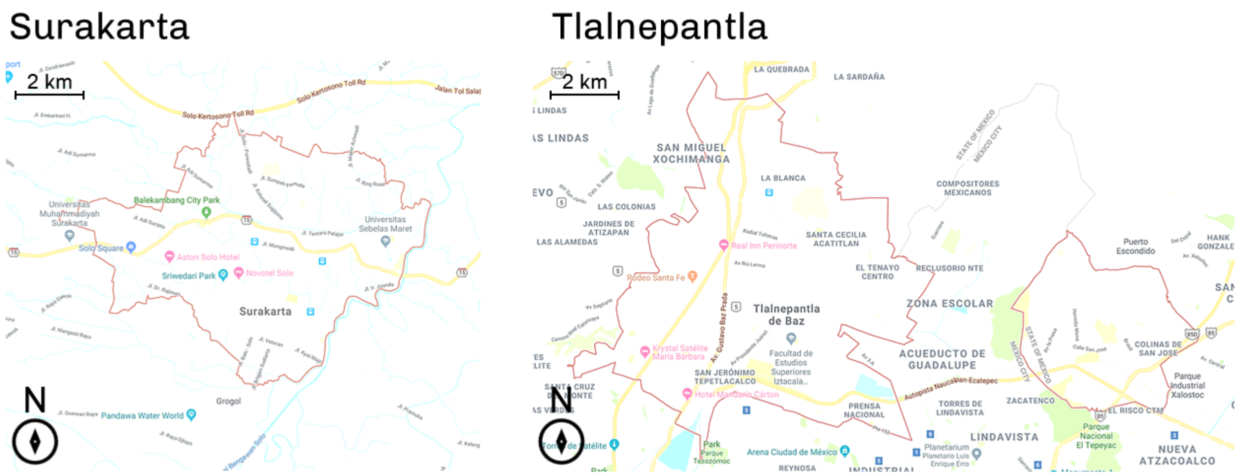

Figure 12 - Maps of Surakarta (left) and Tlalnepantla (right) at the same scale for comparison.

With more than 20 million inhabitants, Greater Mexico City (Zona Metropolitana del Valle de México) is more than three times larger in terms of population than metropolitan Surakarta (Subosukawonosraten) with, almost 6 million inhabitants. Surakarta is the main city of its metropolitan area, while Tlalnepantla is a suburb of the massive Mexico City metropolis. This means that a large number of people commute daily to Mexico City from Tlalnepantla, while Surakarta must have infrastructure to accommodate the people that commute into it every day.

In terms of public transportation, Tlalnepantla is served by the Tlalnepantla station of the Tren Suburbano commuter rail, the Tenayuca station of the Metrobus Mexico City bus rapid transit system (BRT) and a plethora of semi-formal bus routes managed by small businesses. Most of the connections in Tlalnepantla serve to connect the city with the larger metropolitan area, particularly to Mexico City proper. On its side, Surakarta is served by the Adisumarmo International Airport and four train stations for the national railway company, Kereta Api Indonesia. The train stations provide long distance services to travel to other parts of Java and for commuters around Surakarta. Public bus transportation is served by two BRT lines for the Batik Solo Trans (BTS) system and by a network of semi-formal buses around the city.

In Tlalnepantla private transportation consists mostly of private cars, while in Surakarta it is mostly motor scooters (sepeda motor), with car ownership on the rise. Both cities have access to Uber and local competitors for mobile ride hailing. Surakarta also has mobile ride hailing for motor scooters.

At the national level, smartphone penetration in 2016 was at 21% in Indonesia and 35% in Mexico. It is safe to assume that these levels were higher by the time of the experiments in 2018. The percentage of individuals using the internet in both places is very different, with Mexico having a rate almost twice that of Indonesia, 64% versus 32%. Unfortunately, there is no up-to-date data for both cities about internet usage and smartphone penetration.

Table 1 - Comparison of different socioeconomic indicators for Surakarta and Tlalnepantla

	Surakarta (Solo)	**Tlalnepantla**
Role in metropolitan area	Main city	Suburb
Population	499,337 (2010 Census)	664,225 (2010 Census)
Area (square kilometers)	46.01	83.48
Density (people per square kilometer)	10,853	7,957
Part of	Central Java (Province)	Estado de Mexico (State)
HDI (state or province)	0.705 (Central Java, 2017) (Badan Pusat Statistik (Statistics Indonesia))	0.745 (Estado de México) (Programa de las Naciones Unidas para el Desarrollo en México, 2015)
Metropolitan area name	Subosukawonosraten (Badan Perencanaan Pembangunan Nasional, 2009)	Zona Metropolitana del Valle de México
Metropolitan area population	5,984,519 (2010 Census)	21,157,000 (2016 estimation)
Smartphone penetration (national,2015) (Poushter, 2016)	21%	35%
Percentage of Individuals using the Internet (national, 2017) (International Telecomunications Union, 2018)	32.3%	63.9%

In terms of the Human Development Index (HDI), both Central Java and State of Mexico sit within the “High” HDI category (within which values range from 0.7 to 0.8, according to the United Nations Development Programme).

As can be seen in this section and in Table 1, both places are quite different because of their geography and government structures, but they also share many of the common conditions that urban communities in developing countries face, such as: fast-paced urbanization, informal housing and public transportation, and a smartphone market and internet communications that are not yet saturated.

4.2.2 Partner Organizations

Planeación y Desarrollo is an organization based in Mexico City, led by Angélica Garnica Sosa. At the time of the research, the organization was scaling down as funding was scarce and many members were changing their job positions to work in the government of Tlalnepantla. They have produced work that focuses on local development, land use planning, social strategies for violence prevention, and safe urban mobility (Planeación y Desarrolo S.C., 2018).

Kota Kita is an organization based in Surakarta, led by Ahmad Rifai. During the development of this research, the organization had about 12 full time workers. They focus on advocating for increased participation of communities in solving urban issues and have worked on participatory budgets, climate change risk assessments, cycling advocacy, and creating a local urban data portal (Kota Kita, 2018).

Both organizations had previous relationships with me and the Flocktracker development team. They were already aware of the capabilities of the software and had hands-on experience with it in some data collection projects led by their respective teams.

4.3 Experimental Procedures

I coordinated with the two NGOs described above to conduct a total of four experiments. The process started by designing and developing data collection projects for the Flocktracker platform, one with each organization. I inquired with program managers about what initiatives they were working on and if there was a perceived need in the organization for more in-field data about a subject related to their initiatives. Planeación y Desarrollo's members were, at that moment, developing initiatives for public safety in Tlalnepantla while

Kota Kita was working with community members on improving cycling conditions for women in their program Women on Wheels. These subjects were selected for Flocktracker projects. I facilitated access to the Flocktracker tool for program managers at the NGOs and participated in the process of creating one data collection project in Flocktracker for each organization. This included technical advice about the tool, sharing previous experiences with in-field data collection and community mapping, and designing the contents of the project itself (i.e., the questions that the survey in the Flocktracker project would contain). The next chapter provides a more detailed description of the data collection projects' design, content, and structure.

After the data collection projects were designed with each NGO, for each experiment I assisted the NGO in carrying out these steps:

- Implementing engagement strategies for recruiting community members to participate in the project, including:
    - In person invitations to professors and students in local universities,
    - Canvasing at public events with leaflets,
    - Online advertisement campaigns on social media, and
    - Invitations to attendees to other community events;
- Planning and hosting a community event at which
    - Participants filled out pre-surveys about their perception of urban data and open data,
    - NGO leaders and I taught participants about the project,
    - I instructed participants on how to set up and use Flocktracker, and
    - We conducted hands-on data collection with the participants; and

- Facilitating participation after the community event by:
    - Inviting participants to collect data for two weeks, and
    - Requesting, after two weeks, that participants respond to a post-survey intended to measure changes in participants' perception of the usefulness of data collection.

This approach of giving community members an initial experience with the data collection tool is treated as the seed in the "democratization model" described in the previous chapter.

Although these steps describe the original setup of the experiment cases, the strategies for the outreach, the community event, and the data collection experience itself changed in the ensuing cases according to the perceived needs that each NGO stated, in terms of: improving the potential for engaging community members, and considering time constraints, human resources at the organizations, and their interest in undertaking a new experiment case. This led to a single experiment case in Tlalnepantla and three separate cases in Surakarta. I also informed both organizations about each other's approaches. I describe these changes in the descriptions of each case in the next chapter.

4.4 Survey Instruments

I designed the pre- and post-surveys to measure changes in participants' perceptions of the usefulness of urban data collection and open data. In the first two experiment cases, the pre-surveys were administered via pen and paper and the post-surveys were administered by email. In the last two experiment cases, they were administered within the Flocktracker app when new users joined the platform. A unique identifier associated each user with their Flocktracker usage log, described in the next section. The pre-experiment survey included questions about an individual's gender, age, educational attainment, and knowledge of other software, to gather a general sense of the individual's "tech savviness." In addition, the survey asked participants their opinions about data, including:

- The role of citizens in data collection, the availability of data to citizens, and the importance of data to citizens for understanding their communities;
- The types of data (e.g., about business, crime, environment and weather, etc.) an individual prefers be made available in their communities; and,
- Perceived challenges to using data in local governance.

Two weeks after the initial community event, I sent a follow-up survey to participants, through the phone app, that repeated the entire set of questions in the pre-survey. This method intended to measure any changes in the participants' opinions after they used the app for the community data collection project and thereby provide some evidence of the project's impact. The survey instruments are presented in Annex 1 – Survey instruments.

4.5 Instrumentation of the platform

The server platform was instrumented to allow for tracking the usage of the different resources the server provides to understand the usage of these resources in a granular manner and to observe the spread of the use of the tool in a community (Figure 13). This makes it possible to assess the "democratization theory," to see if the seed project entices some participants to create and join other projects, to characterize different aspects of the seed projects conducted and to observe a potential network of community data collectors arising from these seeds. In combination with the pre and post-surveys, this instrumentation allows us to investigate if the amount and types of data collected are correlated with perceptions about the value of data or prior knowledge of other software (i.e., technical capacity). The usage tracking was implemented on the server to limit the complexity and resource expenditures of the Android app and to limit the amount of data the app had to send to the server.

Figure 13 shows an example of the timestamped data recorded in the server The columns contain, from left to right: the time of access, the accumulated total calls to any endpoint in the server, total successful calls and total failed calls. Each row represents the time any of the variables in the columns are updated, giving a detailed time series of the behavior of the platform's usage.

user_snapshot_time	user_total_calls_to_server	user_total_successful_calls_to_server	user_total_failed_calls_to_server
2018-04-28 16:58:20+00	21	19	2
2018-04-29 05:37:21+00	56	33	23
2018-04-28 17:01:46+00	35	33	2
2018-04-28 17:03:51+00	38	36	2
2018-04-28 17:32:11+00	50	50	0
2018-04-28 17:32:12+00	14	14	0
2018-04-28 17:06:09+00	41	39	2
2018-04-28 17:32:12+00	41	39	2
2018-04-28 17:30:02+00	6	6	0
2018-04-28 17:12:55+00	23	21	2
2018-04-28 17:38:02+00	33	33	0
2018-04-29 05:38:55+00	57	33	24
2018-04-28 17:33:30+00	51	51	0

Figure 13 - Example of the user data collected on the server.

Recording on the server rather than on the app also helps to limit potential data leaks since the only communications to the server are directly originated by user interactions with the community mapping functions and no other recording or storage of activity happens on the app (client) side. For example, the app does record the current location and time when a user uploads a survey, but it is never aware of how many times the user has downloaded all the surveys, how many times she made changes to a project, or how many times the user uploaded surveys to different projects, among many other behavioral traces that are recorded on the server side. Only the research team had control and access to the server-side usage database, which was not shared with anyone outside of the research team and is not available to the partner NGOs.

The variables were tracked every time a user made a request to a resource endpoint in the server by using the Android app while connected to the internet. An endpoint is a unique address (Uniform Resource Locator, URL) to a function the server is capable of providing; for example, the “upload a survey” resource allows the app to request that the server receive a new survey and its related metadata, including survey answers, location data, time, and pictures. If the server accepts the request, the data transfer process starts, and the server sends a confirmation of the successful upload when it is finished saving the survey data. When an endpoint is accessed, all the variables related to it are updated and stored in a master table.

Three different categories of variables were recorded:

- Server – These variables are recorded globally, aggregating the usage of all the users and projects in the platform. The variables are mostly to assess the health of the complete system. This can be done, for example, by examining the rate between successful and unsuccessful calls to the server.
- Project – Related variables are categorized by projects in the platform and allow for the identification of levels of usage pertaining to a specific project. With the timestamps recorded, the “pulse” of a project can be studied by using time series charts and related analysis to investigate how interactions with a project change over time. For example, if

people add more participants to a project when the community mapping is just starting or if they do so when some data is already collected. This could allow for testing for correlation of changes of engagement strategies and participation amounts over time, for example.

- User – User-level variables allow understanding how individual users interact with the different resources on the server (projects, maps, data download, user management). Each time a user accesses an endpoint, a tally of the user´s access to that endpoint is updated. The last time of access is recorded as well. This allows testing for correlations between users' data uploads and joining or creating other projects, for example.

The full list of variables tracked can be found in Annex 2 – List of variables for server-side instrumentation for usage tracking. Combined with the actual data collected by users in the field, all the analytics proposed in the instrumentation of the tool could help to bring a clearer picture of the spread of the tool and the user interactions that happen on it.

4.6 Assessing the role of different local partnerships and outreach strategies

Finally, I attempted to evaluate the effectiveness of each of the different partnerships and the different outreach strategies used in the different experiment cases, trying to understand how they might have a more positive impact on the spread of data collection in a community. For the partnerships observed in this research, aside from the NGOs, local transportation authorities, universities and city governments were engaged and invited to be part of the interactions with the participants to create the community mapping projects. I attempt to evaluate the role these institutions played in participating in the experiments and the impacts of that role in the different activities of the experiment. Based on the experience, I also try to identify potential pitfalls in the engagement process with the institutions when inviting them to the experiment and when informing citizen participants about their participation.

In the case of the outreach techniques, I evaluate them in terms of the amount of people they might attract to participate at different levels – from being informed of the community data collection project, to attending to the main event, to continuously collecting data for the project after the initial event – and for their ability to entice people to generate their own data collection projects.

These partnerships and outreaching strategies were chosen by the NGOs given their own resources and timing requirements, so the experiments do not isolate the effects of each of the partnership or strategy. My observations and conclusions drawn from studying these elements provide an initial understanding of their strengths and weaknesses and how they might be utilized in the future.

4.7 Effects of researcher participation in the co-design and implementation of experiments

It is important to emphasize that my participation in each experiment informed my evaluation and subsequent conclusions. I participated in the process, taking the role of a facilitator and technology advisor when needed. In my direct involvement in creating the strategies for engagement and helping to develop the data collection projects for the Flocktracker app, I always leaned towards giving the NGOs space for argumentation and ultimate decision-making power about the content of the data collection process and the engagement strategies. Although I advocated for contacting local partners to give support for the projects, the ultimate selection of these partners was done by the NGO project managers. As per the request of the organizations, I gave them updates about the strategies the other organization were taking. Each experiment was informed by the previous experiments.

My prior relationship with the organizations through their use of the Flocktracker platform presents the possibility for bias in the conclusions of this thesis. For example, the organizations might not have selected Flocktracker as their platform for collecting data without my input. In the case of Surakarta, where three different experiment cases were undertaken, the earlier experiments may well have influenced the later ones, since some of the participants in one experiment case might participate in the subsequent ones. Nonetheless, this seems rather unlikely in these cases, due to a very low level of observed participation and interest in collecting data.

5 Experiments

Four different experiment cases were setup and deployed in the field following the timeline in Figure 14, which shows the start and end of the intended data collection period with members of the communities. Following the democratization model outlined in Chapter 3, they are all meant to be separate “seed” projects to better understand how digital community mapping might serve as a niche for kickstarting the widespread use of digital data collection tools in locations of the developing world. Three experiments were set up with Kota Kita and one with Planeación y Desarrollo. The engagement strategies evolved as the lessons from one experiment were passed on to the planning of the next. The data collection project was unique for each of the NGOs and did not change in the different experiments.

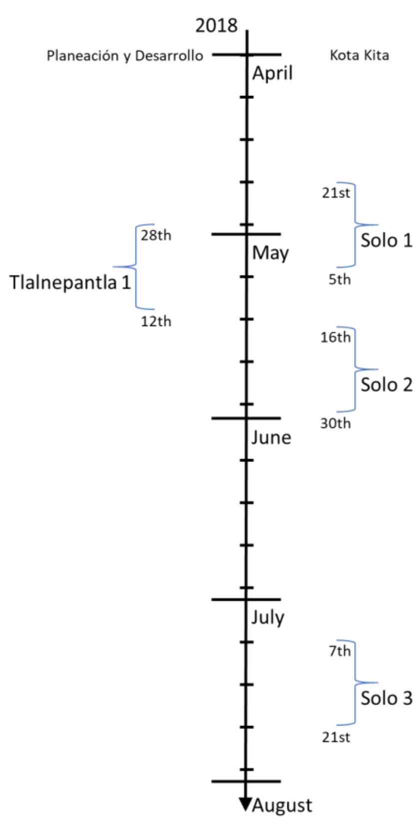

Figure 14 - Timeline of experiments in Mexico and Indonesia.

5.1 Community data collection project design and content

A data collection project was setup with Planeación y Desarrollo´s team, and a different one with Kota Kita's. The projects were developed in parallel with both organizations. The tool was provided to them and I served as a technical consultant on the use of Flocktracker and the design of the data collection process. In both cases, the NGO members were able to setup the projects by themselves. Projects were also tested in the field by members of the organizations and myself before the experiments were started.

The projects were designed with the goal of collecting data that would allow for the respective NGOs to further their advocacy work and to allow for a very simple user experience for the participating community members. The subjects were selected because the NGOs' deemed them interesting for their agendas and their communities.

The general structure (Figure 15) of the projects consists of the following components:

a) Report classification – A question asks the user to classify what they are observing by selecting from among different categories (e.g., street cleanliness).
b) Report subclassification – After the initial classification, the next question prompts the user to sub-classify the report within the initial class and to add more data relevant to the class. This took advantage of the branching capabilities of FlockSON.
c) Rating – The user is asked to share rank their feeling/perception about the reported item, on a Likert Scale from 1 to 5 stars (★).
d) Picture – The user can take a picture of the observation.
e) Comments – The user is provided the opportunity to add any additional comments about the report.

The expected method of participation was that community members would report what they observe in the public space using the project setup on their smartphones.

The structural similitude of both projects could allow for one to one comparison of the amount of data collected and the qualitative aspects of the survey. Both projects share the same sections except for the subclassification one. It is important to mention that even though the Planeación y Desarollo project had a total of 12 questions, the participant would only see 5 or 6 questions per report including the picture, depending on the content of the report. This was possible because of the branching of the survey based on the classification answers (Figure 15). As such, the survey developed for Planeación y Desarollo only needed five or six questions per report while that for Kota Kita always needed four questions. This was an intentional design feature to allow for both easy understanding of the process and quick data collection in the field.

The Planeación y Desarrollo project was developed to help deepen the NGO´s understanding of the public space conditions that affect safety perceptions in Tlalnepantla. With Kota Kita, a project named Pedalista, related with their Women on Wheels cycling advocacy initiative, was designed, allowing for people to report on the conditions of the roads and other infrastructure related to bicycling in Surakarta.

Figure 15 shows a flow chart of the projects developed with each NGO. Both share most of their sections but the subclassification one. It is important to mention that even if 1. has a total of 12 questions, the participant would only see 5 or 6 questions per report including the picture, depending the content of the report. The survey developed in 1. only needs 5 or 6 questions per report while 2. always needs 4 questions. This was an intentional design feature to allow for both easy understanding of the process and a quick data collection when in the field.

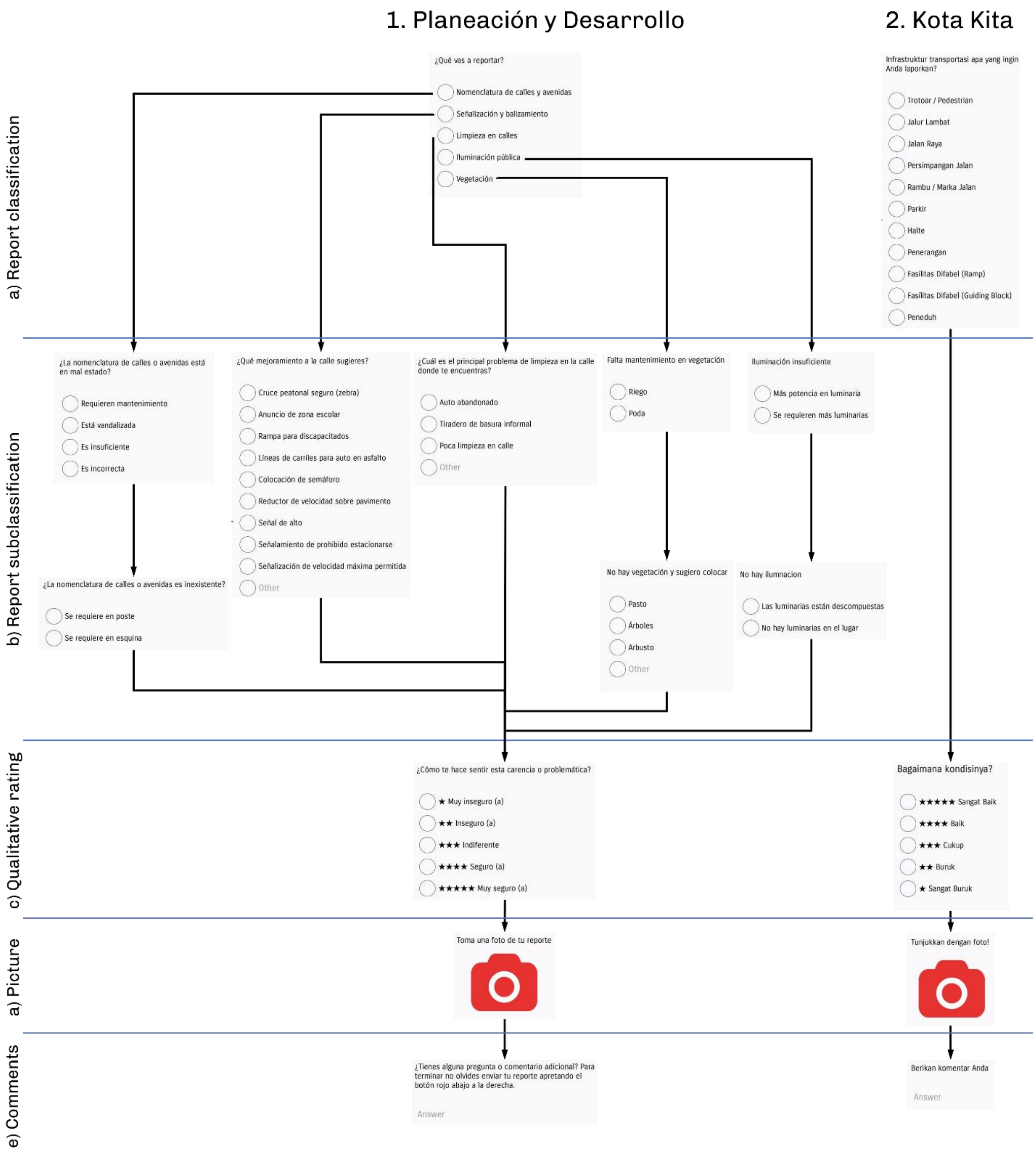

Figure 15 - Flow diagram of the projects developed with Planeación y Desarrollo (1.) and Kota Kita (2.).

5.2 Description of the experiment cases

This section describes the four experiments in chronological order (Figure 14). Each case description includes five dimensions: Partnerships, community outreach, interaction materials, community event, and post event participation. For each case description, the outcomes subsection discusses the results, learnings, and decisions taken for the deployment of the next experiment.

5.2.1 Solo 1

This was the first experiment case in the series. The main community event was held in the facilities of the Surakarta City Transportation Agency (Dinas Perhubungan Kota Surakarta) on April 21st of 2018. The data collection period was expected to run from this day until May 5th.

5.2.1.1 Partnerships

The Surakarta City Transportation Agency, the local transit agency, was the key partner. They provided the venue for a community meeting they organized jointly with Kota Kita to educate women in Surakarta about cycling safety under a program called Women on Wheels (Figure 16). The Flocktracker community event was held immediately after this event.

Figure 16 - Community demonstration of bicycle safety in the Women on Wheels event organized by Kota Kita and the Department of Transportation of Surakarta.

5.2.1.2 Community outreach

Outreach in this case was done through the partnership with the Surakarta City Transportation Agency. The Women on Wheels program, for outreach and education about road safety for women in Surakarta, had already planned the Saturday April 21st event (Figure 16). Women participants were already planning to come to the event and the Flocktracker/Pedalista project was a non-announced addition to the schedule of activities for the day. From initial conversations with Transportation Agency members and Women on Wheels participants, it was expected that, due to their attending the event, the attendees would be interested in the data collection project because of their interest in transportation in Surakarta.

5.2.1.3 Interaction Materials

For the event, Kota Kita and I prepared a presentation to teach people about the Pedalista project. The presentation included the following sections:

- What are urban data and why do we need it?
- What are open data?
- What is Flocktracker?
- What is Pedalista?
- How to setup Flocktracker and the project named Pedalista?

5.2.1.4 Community event

About 60 women 50 years old and older attended (Figure 17). As mentioned, the Pedalista/Flocktracker collaboration took place just after a road safety event, during which attendees were taught about safe ways to navigate Surakarta's roads. After the aforementioned short training presentation (about urban data, Flocktracker, etc.) (Figure 18), we expected to be able to setup the app on the participants' phones and collect some data points near the venue with them. Although we prepared a computer slideshow presentation to guide the participants on the process, all of them needed direct one-to-one assistance to work with the app.

Figure 17 - Participants in the Women on Wheels community event.

Figure 18 - Titis Efrindu and me presenting on the event.

Many participants had difficulties understanding how to setup Flocktracker on their Android phones. Informally, we were informed that most of the women have very limited experience with smartphone technology and that, even if they have smartphones, they tend to only use them for basic messaging, picture sharing and phone calls; younger family members tend to be the ones setting up their phones and apps for them.

Even as members of Kota Kita and I helped participants setup the app on their phones, trying to be as simple with our explanations as possible (Figure 19), many of them lacked basic understanding of how internet services work. For example, one woman asked what a password is and asked one of the Kota Kita members if she could tell her what her password is to setup an account. This makes it clear that she did not have a concept of what is a password, who should know it, and the importance of keeping it secret. The Flocktracker user experience was clearly not easy enough for this group of participants.

Figure 19 - Woman signing up for Flocktracker in her Android smartphone.

The user experience problems were aggravated by a failure on the server side of the platform. The error meant that every few interactions, the server would go into an irreparable error state that required a full restart of the server to return it to normal functioning. The combination of server errors, limited ability to support users directly, and the app's user experience meant that all the participants became uninterested in the process very quickly and left the venue after 15 minutes of presentation and 15 minutes of trying to setup the app on their phones.

5.2.1.5 Post event participation

No data collection was observed after the event. This is not unexpected given the very low understanding of the app that the users shown when presented with it.

5.2.1.6 Outcomes – Technology

From this experiment, the outcomes were twofold. On the community engagement side, having people from only one demographic, mostly women 50 years and older women, hindered the development of the workshop. Even if some of the participants did understand how the app and the project worked during the event, it was impossible for six people (five members of Kota Kita and me) to setup the app and explain how it works directly to all of them without losing interest of the group.

On the other hand, the server side of the app failed on the day of deployment. Though eventually restored later, the delay made it impossible to keep the attention of the attendees during the workshop. Most of them left within the first twenty minutes when the app did not work. The failure of the server helped to identify some bugs when signing up to the platform. All of them were subsequently resolved and did not present problems in subsequent experiments.

One challenge when creating the data collection project was the lack of convergence between the communities' tech savviness and the anticipated user experience for the platform. Flocktracker's design was based on the expectation that users would be able to use basic social media apps, which was not the case for this community of participants. In retrospect, to allow an audience with no knowledge about how to use smartphones to participate, we could have setup an app that enables uploading mapping points completely anonymously, with no login information needed, but this would also hamper the ability of the organization deploying the project to manage it, such as limiting access by different people to different projects. At a larger scale, a completely self-guided process that does not need an in-person explanation about the use of the platform would require such a friendlier system to facilitate non-directed participation of the less tech savvy members of a community. In the subsequent experiments, a leaflet outlining the steps to collect data was handed out to all participants. We also ensured that outreach included a more diverse group of participants, to increase the likelihood of people knowing how to operate the app and also inviting these individuals to help the ones that did not know how to set the app up.

5.2.1.7 Outcomes – Participation process

Participants in the event were already interested, to some extent, in road safety, but were not necessarily looking for technological solutions to their concerns. This case suggests that a multi-tier process that starts with collecting data, but that might or might not have real consequences on participants' daily lives, apparently had very little perceived value for participants (regarding their presumed interest in improving their road safety now). In a workshop with the Kota Kita team after the event, we concluded that the presentation we used to try to explain issues of transparency, Open Data, and community mapping was too complex for most audience members to grasp, in a single event that also included community data collection.

5.2.2 Tlalnepantla 1

Figure 20 - Angélica Garnica presenting on the Cultura de Paz, Redes Vecinales community event before the Flocktracker event.

The second experiment case in the series, in Tlalnepantla, Estado de Mexico, with the collaboration of Planeación y Desarrollo, focused on the San Miguel Chalma neighborhood. Given the then current downsizing of personnel and funding in Planeación y Desarrollo, their team´s participation in the outreach efforts was very limited until the week immediately before the event. The initial plan included a standalone event for the community mapping, but this was eventually set up as part of a community outreach event of the Tlalnepantla municipality's Cultura de Paz (Peace Culture) program, focused on Redes Vecinales ("Neighborhood Networks"), an integrated effort to build local networks and technology for improving local safety. The event involved the Municipal President Denisse Ugalde Alegria presenting the Municipality's safety prevention efforts, together with other government officials, including the chief of police, and the developers of the Redes Vecinales program. Angélica Garnica, the CEO of Planeación y Desarrollo and Titular del Área Integradora de Cultura de Paz (Incumbent of the Peace Culture Integration Area) attended the event and requested the space for Flocktracker to be part of it (Figure 20). The event was held on April 28th of 2018. The data collection period was expected to run from this day until May 12th.

5.2.2.1 Partnerships

We invited students from a Land Use Economics class from the Facultad de Estudios Superiores of the Universidad Nacional Autónoma de México-Campus Acatlán. The professor of the class encouraged his students to participate and four of them attended. The Flocktracker community mapping event was held after the Cultura de Paz program event.

5.2.2.2 Community outreaching

The initial goal was to reach 50 to 100 attendees. We took active measures to invite younger people (based on learning from the first Solo experiment), who we expected would more easily understand how to setup the app. We used four outreach channels: social media posts, personal interaction in local universities, pasted posters in the same universities, and an invitation to participate issued during the Redes Vecinales event itself. The final date and location of the event was not defined by Planeación y Desarrollo until four days before the event, which limited the number of people that could be contacted and invited to the event (outreach could only start after the time and location were defined). Some of the people that expressed interest in the event through our personal outreach efforts mentioned that they could not attend due to existing conflicts. Some even asked about other potential events, so they could participate.

Figure 21 - Poster used for the Tlalnepantla event Designed by Ilse Jiménez Molina, Ali Al-Sammarraie and me.

Two members of Planeación y Desarollo (IIse Jiménez and Ross Jiménez) and I conducted field outreach in Tlalnepantla from April 25th to 27th, visiting all the large academic institutions in the area:

- Universidad Nacional Autónoma de México – Facultad de Estudios Superiores Iztacala;
- Universidad Nacional Autónoma de México – Facultad de Estudios Superiores Acatlán; and

- Instituto Tecnológico de Tlalnepantla.

We invited people from the Urban Planning, Architecture, Sociology and Geography departments to participate in the event. We also posted dozens of posters (Figure 21) inviting people to participate in the event. The UNAM campuses were more open to allowing the posters in their facilities than the Instituto Tecnológico de Tlalnepantla, where permission was required. We also asked their outreach departments if they could send emails to their students about the event. Administrative personnel did not approve of this email outreach because the timing did not allow for complying with the required institutional approval procedures. As mentioned above, we also presented to and invited students from a Land Use Economics class at the Universidad Nacional Autónoma de Mexico-Acatlán. About eight of them came to the day of the event.

The same poster used physically at the universities (Figure 21) was disseminated electronically through the Planeación y Desarrollo Facebook page and my personal one. We observed that some colleagues and friends shared this. Four people that attended the event learned about it from these social media posts. Other people contacted us to know if they could join, but they were from other areas of the Mexico City Metropolitan Area and Tlalnepantla was too far for them to join. A WhatsApp group chat was setup with the students and online connected people to coordinate their attendance to the event.

5.2.2.3 Interaction Materials

Aside from a more streamlined presentation than the one used in the first Surakarta case, a leaflet was prepared and made available to participants (Figure 22). This provided people with a copy of the instructions of how to setup and use the app on their mobile devices and allowed them to share the information about the project with their friends and families after the event.

The leaflet contained:

- A summary of the objectives of the event (collecting data to improve public safety in Tlalnepantla),
- An invitation to participate to anyone that lives or spends time in Tlalnepantla,
- Instructions to download the app from the Google Play Store,
- Instructions to sign up and join the project,
- Instructions to collect data,
- Basic instructions to visualize the data, and
- Web links to share this information with others.

Figure 22 - Leaflet to teach participants how to setup the app and the Tlalnepantla project.

The leaflet proved to be very useful in enabling younger participants to use the app. All the participants who were students or professionals were able to setup the app using only the leaflet as a guide. It took each of them less than five minutes to set up the app on their phones and be ready to collect data.

5.2.2.4 Community event

On the day of the event, we invited the students and others who signed up online to meet us in Tlalnepantla Centro one hour and a half before the Tlalnepantla municipality event. The purpose was to set up Flocktracker ahead of time with them and ask them for their help with setting up Flocktracker on other participants' phones (including older people and others who might have problems understanding how the app works) during the community event. This also provided the opportunity to test if the users were able to setup the app by themselves with only the leaflet. All of them were able to set up the app with no issue in only a few minutes. We then travelled to San Miguel Chalma for the main event.

Figure 23 - Students helping older participants to setup the app on their phones.

After the Redes Vecinales event finished, participants were invited to stay for the Flocktracker event. A group of four women more than 50 years old decided to stay. One of them subsequently did not participate because she did not want to share her information in the initial survey. She mentioned a history of government retaliation against people that expressed themselves in this kind of surveys as a reason. The yournger people that had setup their app earlier in Tlalnepantla Centro helped the remaining older participants to setup the app on their phones (Figure 23). When all the people had answered the initial survey and setup their phones, Ilse Molina and I gave the Flocktracker presentation (Figure 24).

Figure 24 - Community meeting in Tlalnepantla.

After the presentation, the 14 participants were divided into three teams to collect data around the venue area (Figure 25). We walked for 30 minutes around the neighborhood, using the app to collect relevant observations. Younger participants helped the older ones as needed. After this field data collection, we held a debriefing session in the original venue to talk about the experiences. During this debrief, participants indicated that the event helped increase their awareness of the need for evidence gathering and community participation with authorities to start making changes to, and improvements in, the safety of their community. They also mentioned that they are so used to not having any feedback from the government that most of the time they feel like giving up on efforts to collaborate with them.

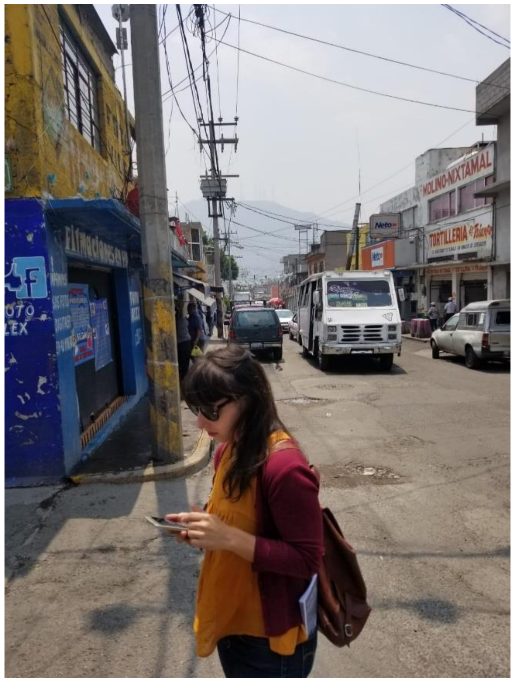

Figure 25 - Participant Collecting data in the street using Flocktracker.

5.2.2.5 Post event participation

After the event took place, no additional data collection was observed. Several possible explanations exist. First, the students and young professionals that attended, even if they expressed interest in participating later, did not live in Tlalnepantla, where the data collection efforts were to be focused. Another potential issue is that the San Miguel Chalma neighborhood is perceived as particularly unsafe, even in the Tlalnepantla context. This might have inhibited people's willingness to take out their phones in public and collect data when alone. All the participants from the neighborhood were women, such that their personal safety risks could have been higher than for potential male participants. They did express their willingness to share their experiences and the leaflet with instructions to their family members and friends, to encourage their participation.

A follow-up event to invite people to see the results of the data collected was not organized, primarily because of Planeación y Desarrollo's lack of resources, as mentioned earlier. The lack of further participation could have been caused by missing follow up goals from the involved organizations to address the problems highlighted by the data collected and/or the lack of a closure event after the intended two-week data collection period.

5.2.2.6 Outcomes – Technology

As mentioned earlier, the technology worked well for the participants in the event. Having younger people who better understood the interfaces helping older people with less familiarity with smartphone apps was key to allowing everyone in the event to actively participate in the data collection activities and share their observations.

5.2.2.7 Outcomes – Participation process

Even if participation was low compared to our original outreach goals of attracting 50 to 100 people, those few who did participate described the experience as positive. Informally, they mentioned that the exercise led to their improved understanding of the value of data and evidence-gathering for demanding better services from authorities. They mentioned this as a component of "doing their part' to improve their community. Except for the previously mentioned case of the older woman who expressed concerns about participating because she was afraid of government retaliation if she complained about the local conditions, the rest of the group mentioned that they would like to collaborate more with the government to improve their place.

Of the 14 participants, 10 were students and some City Planning professionals that saw the outreach call through our Facebook posts. Most of them came from of other parts of the Metropolitan Area of Mexico City and only two of them were from Tlalnepantla proper. When asked about their willingness to travel across Mexico City, in some cases for more than an hour and a half, to participate in this process, they mentioned that they were interested to learn about digital data collection technologies and outreach methods and that they see data collection as a big part of participative cities. This might be indicative of the potential of similar tools for digital participation to increase the likelihood of younger people to participate in community-oriented events.

Figure 26 - Screenshot of the data collected in Tlalnepantla showing a color-coded view of the types of data collected.

Unfortunately, the participation of Planeación y Desarrollo in the research could not continue after this experiment because of two main factors. One, as already mentioned, their operations were being spread thin given lack of funding; most of the members of the organization were not working full-time by the time of my initial contact with them in December 2017. By the time the experiment was undertaken, the leading team of the organization had completely moved to work on the Cultura de Paz program at the Tlalnepantla municipal government (which led to the research becoming part of the Tlalnepantla government's efforts to outreach to their community). In addition, the timing of the project coincided with the campaign process for the 2018 general elections in Mexico. By law in Mexico, government organizations cannot conduct outreach to their communities during election season since this could be interpreted as partisan campaigning.

5.2.3 Solo 2

This was the second experiment attempted in Surakarta, which built from the experiences from the previous two cases to improve the process. In this case, all the participants were students, specifically members of the Universitas Muhammadiyah Surakarta (UMS). In contrast to the previous two experiments, in this case I did not participate on the ground in any of the in-person outreach, training, or field data collection activities. I did participate in the design of the outreach materials and the slideshow presentation content and gave support to Kota Kita live during the time of the event via phone calls and instant messaging.

5.2.3.1 Partnerships

This experiment also counted on the partnership with the Surakarta transportation agency, but only in promotional media and in the presentation materials of the digital community mapping event. Professors of social science at the UMS invited their students to the digital community event held at the university.

5.2.3.2 Community outreach

Kota Kita staff members, Titis Efrindu and Icha Tamrin, conducted the community outreach by contacting USM professors who invited them to give a presentation at the university about the Pedalista project. During this outreach, the professors expressed their interest in sharing the use of Flocktracker with their students and inviting the students them to participate in the project.

5.2.3.3 Interaction Materials

The presentation given by Kota Kita staff to the USM students was a shorter version of the one used in the first Solo experiment. It focused more on giving attendees an overview of Flocktracker, the project (Kota Kita's Pedalista), and how to install the app and start collecting data.

5.2.3.4 Community event

The community event was held at UMS on May 16th, 2018. About 50 undergraduate students of City Planning were invited to the abovementioned presentation. The presentation took place in the school auditorium after which the students were supposed to undertake a field data collection experience in the area around the university. The latter field data collection effort did not, however, ultimately happen because of time constraints. Students and professors had to go to other classes after the presentation. In the end, 34 participants at the event installed the app. Aside from some problems on the server side that made the installation process slower for some users, no major challenges were perceived when teaching the students how to use the app.

Figure 27 - Students in Universitas Muhammadiyah Surakarta installing Flocktracker on their smartphones.

5.2.3.5 Post event participation

No data were observed to be collected by participants after the initial event.

5.2.3.6 Outcomes – Technology

Since all the participants were young and educated, the technology was easy to understand and setup by all the participants. A couple of errors specific to some Android devices that did not allow for usage were found and corrected in the days following the main event.

5.2.3.7 Outcomes – Participation Process

As mentioned, about 50 students engaged in the presentation. A debriefing with Kota Kita suggested that the event might have been misleading for the students. The professors apparently invited the students mostly because they wanted them to learn about Flocktracker as a data collection platform rather than actually participate in the Kota Kita initiative to collaborate on creating the Pedalista map. In other words, expectations were different. This may be representative of a general challenge faced by similar technology-based community interactions: some participants might care more about the technology itself rather than specific topical interest (e.g., traffic safety) of the community organization. That said, regarding the process of seeding the data collection communities, universities and researchers proved to be good partners for reaching large groups of people interested in the technology and savvy enough to use it. In this particular case, balancing the intention of increasing the usage of the Flocktracker platform with using the tool for Kota Kita's direct concern was not well achieved. Finally, as mentioned, the time constraints around the outreach event, which did not allow for the participants to take part in a hands-on data collection experience with Kota Kita, could have also been a factor in why the students did not choose to participate in subsequent data collection.

5.2.4 Solo 3

For this final case, outreach methods were modified to include paid social media advertisements while not including a community event with a presentation as in the previous cases. Instead, people were invited, in-person, to participate during the car free day events that are held every Sunday in Solo. This experiment started on July 7th and ended on July 12st.

Another change relative to the previous experiments involved offering prizes for the people who collected the most data. The awards were to be delivered as data top-ups for the participant's phone. The idea was to motivate more participation in the project. The structure of the prizes for the topmost collectors were as follows (in Indonesian Rupiah with approximate value in USD is in parenthesis):

- 1st place - 500.000 IDR (33 USD);
- 2nd place - 300.000 IDR (20 USD);
- 3rd place - 200.000 (13.5 USD).

The combination of the online paid ads and the contest structure was expected to boost the chances of the project being successful in engaging with Solo's citizens.

5.2.4.1 Partnerships

As in the prior Solo cases, the partnership with the Surakarta transportation agency was maintained, but in this case only in the promotional media used.

5.2.4.2 Community outreach

Outreach was done via social media posts, newspaper mentions, and in-person interactions with community members. In this case, there was no community event focused on Pedalista or Flocktracker. All outreach efforts focused exclusively on inviting people to participate directly in the project rather than to attend a specific event.

On the Kota Kita Facebook page, a paid advertisement campaign was created around the image shown in Figure 28. A total of USD 200 were spent on the campaign which focused on showing the advertisement to Surakarta's Facebook members. The post was seen by more than 33 thousand people and generated interactions with more than 1,500 people. Interactions in Facebook include likes, comments, sharing, and clicking the image on the post.

Due largely to Kota Kita's contacts, the previous two experiments in Surakarta and their interactions with local universities and the local transportation agency, some media interest focused in the Pedalista project. This included at least two mentions in local printed newspapers (Figure 29).

Figure 28-Online advertisement for the Facebook social media campaign.

da Kompromi
Sunday Market

PEMANFAATAN TEKNOLOGI

Yuk, Ikut Rekam Kualitas Infrastruktur Transportasi Kota!

Figure 29 - Mention of the Pedalista project and the Flocktracker app in Solopos, a local newspaper.

Finally, the Kota Kita team attended the Car Free day events organized by the government of Surakarta every Sunday (Figure 30). These events promote active mobility by closing the main roads of the city to cars, thus creating safe spaces for people on foot, bicycles, roller skates, and other active mobility modes. We expected that people in these events would be aware of the challenges of urban mobility and consequently willing to participate.

Figure 30 - Titis Efrindu using the leaflet to invite people to the Pedalista project during the Car Free Day event held every Sunday in Surakarta.

5.2.4.3 Interaction Materials

Materials produced to interact with potential participants are as follows:

- Website – A web page which included information about the contest and how to download the app and start collecting data (https://www.flocktracker.org/pedalista);
- Paid social media ad – A poster (Figure 28) to be included in the Social Media campaign calling for people to participate in the event and including an overview of the contest, a link to the website and basic instructions about downloading the app;

- Leaflet – A streamlined leaflet, allowing for it to be printed on one side of a sheet of paper (Figure 31), which contained information about the key dates of the contest and the prizes.

The leaflet was an improved design from the one used in Tlalnepantla. It explained the sign-up process in a more streamlined manner and also included information about the data collection contest and its deadline. The leaflet was distributed during the personal outreach on the car free day, including to passersby with not time to interact, to guide people about the project (Figure 30).

Figure 31 - Leaflet used to spread information about the third Solo project.

5.2.4.4 Community event

As discussed, there was no community event organized for this case. All personal interactions were one-to-one and in the context of the car free day event held every Sunday in Solo.

5.2.4.5 Participation

Only one person was observed uploading data by herself to the project (Figure 32). We emailed her to learn about her experience and to deliver the award, but she did not respond. Data in Figure 32 is showing surveys filtered by user that collected each data point. Note that only one point was collected by someone that is not a member of the Kota Kita team.

Figure 32 - Screenshot of the Flocktracker app showing the Pedalista project at the end of the contest period

5.2.4.6 Internet interactions performance

The use of a Facebook advertisement campaign allowed for the gathering of some metrics about outreach. They can be compared with the data provided by Google on the Play Store about downloads of the app and ultimately with participation in the project itself. Although the campaign had an original budget of USD200, Facebook only charges for the number of times the advertisement was shown to users (i.e., “impressions”). The final cost of the advertising campaign was USD 66.

Figure 33 - Facebook ad campaign charts provided by Facebook with reach, engagements, cost and impressions during the time of the contest.

Figure 33 shows the performance charts for the Facebook advertisement campaign. The 2-day gap on July 13th and 16th resulted from a problem in processing the original payment. Engagements include likes, reactions, comments, shares and clicks on the advertisement image. Reach refers to the total number of individuals that saw the campaign, while impressions are the number of times the ad was shown to any user. During the two weeks of the intended data collection, the post was seen 88,300 times, reached 33,688 people, and was engaged with 1,556 times. The cost to reach 1,000 people was USD2.00 and each engagement had a cost of USD 0.04, on average.

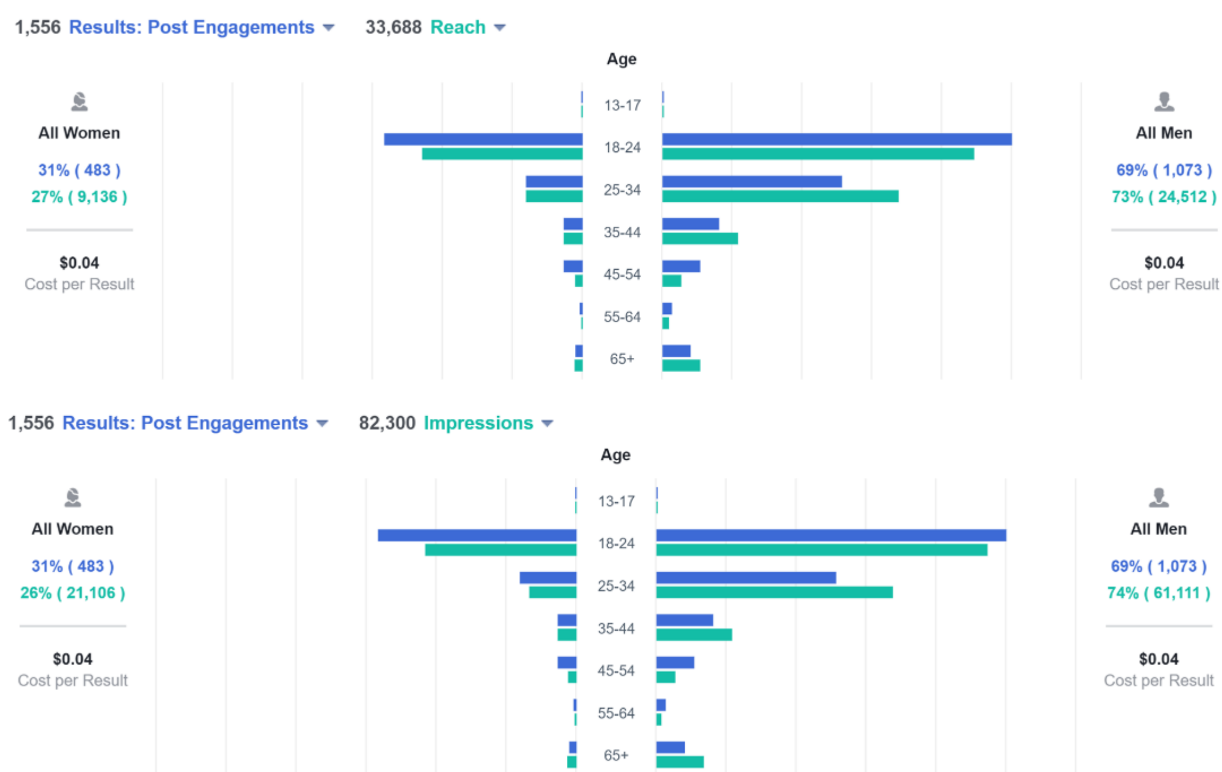

Figure 34 – Gender and age distribution of the people reached by the Facebook advertisement campaign.

Figure 34 displays the gender and age distribution of the distinct metrics provided by the Facebook ad campaign. Overall, the campaign reached more individual men (73% of reach and 74% of impressions), although slightly fewer men participated in engagements (69%). This indicates that women had on average a higher response rate than men after they saw the advertisement. Men had on average 44 interactions per 1000 men reached, while women had 53 interactions per 1000 women reached. All the metrics show a higher number of people in the 18-24 age range, with the 55-64 years age range being the least represented. Facebook reports USD0.04 per impression for males and females. Facebook does not report figures on the gender composition of their users in each place (i.e., number of males and number of females that report living in Solo and are part of the Facebook platform) or the algorithms used to choose who will see each paid advertisement campaign post.

Figure 35 - Pageviews for https://www.flocktracker.org/pedalista/ , in the time of the contest.

The chart in Figure 35 corresponds to views of the web page created for the contest. It was obtained with the Google Analytics plugin installed on the Flocktracker website (flocktracker.org). The web page was viewed a total of 30 times by 27 different users, with a peak of eight views on July 18th. These different viewers could include members of Kota Kita as well as myself. Unique viewers can be counted more than once if they use different devices or web browsers to access the web page. Viewers spent on average 2 minutes and 6 seconds on the web page. Many of the people who reached the site did so through the link available in the original Facebook post.

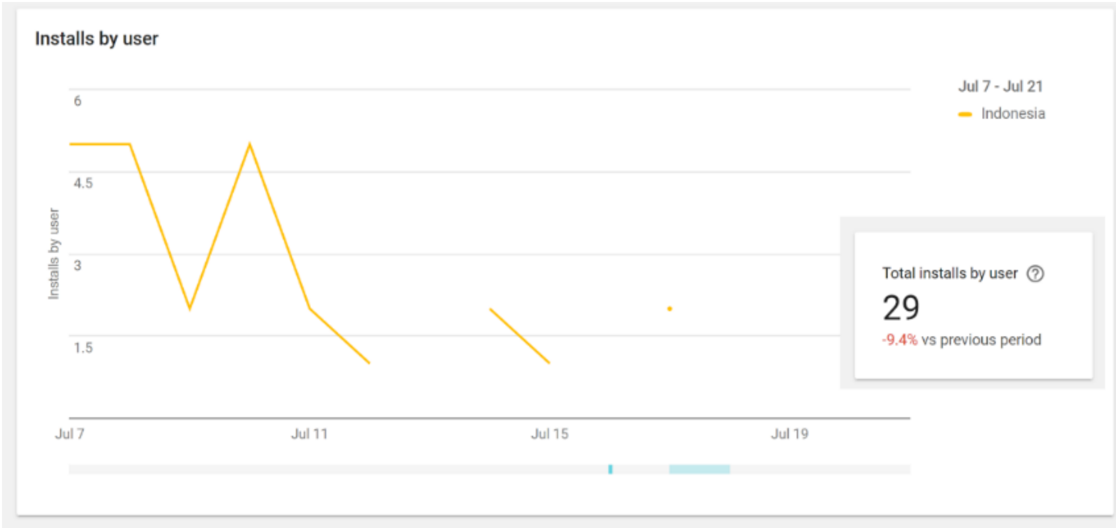

Figure 36 - Daily downloads of the Flocktracker app during the extent of the campaign of the 3rd experiment case in Solo.

Figure 36 shows the number of downloads of the Android app in Indonesia from the Play Store during the campaign. Despite more than a thousand interactions in the Facebook campaign, the app was downloaded only 29 times in total. Downloads were more common at the beginning of the campaign, with a maximum of five downloads per day recorded on July 7th, 8th and 10th, matching the initial days of the outreach and only one of the three car free day weekly events where in-person outreaching was done. No downloads took place after July 15th. Since the app was downloaded by more people than the total amount of viewers of the web page, it is very probable that some people downloaded the app by going directly to the Play Store either from the instructions on the Facebook advertisement or by reading the physical leaflet.

If only the online campaign is considered, that is, disregarding the in-person interactions during the Car Free day and all the preparation work, the one data point collected had a cost of USD66.

5.2.4.7 Outcomes – Technology

Flocktracker was monitored remotely to make sure all of its components were working properly. Only minor layout rendering errors that misplaced some of the interactive graphical user interface components on the screen were identified and corrected as fast as possible. The value of the leaflet and website in terms of effectively providing a short explanation the platform could not be tested in-depth with the community. Participation may have been low because people deemed the setup process as being too complicated.

5.2.4.8 Outcomes – Participation process

Although this experiment was the one that entailed the most work in online and in-person outreach, it lacked the shared communal experiences that come with attending an event, sharing views on the topic, and collecting data together. The high number of interactions via social media indicates some interest in transportation infrastructure improvements, community mapping, and digital data collection, although this also might be related to the announcement of a contest to participate. The very low traffic on the website and low number of downloads of the app, even after the in-person outreach during the car free day weekly events indicate a potential disconnect between what the community finds interesting and what they may actually be willing to participate in.

5.2.5 Summary

This chapter describes the four experiments carried out in two different settings. With respect to the research questions, the cases offer inconclusive evidence about the impact of participating in a digital community mapping experience on people's perceptions about the role of data and community participation. This is because of the very low participation rate in all of the experiments and the subsequent lack of post-survey completion. In terms of the "democratization model," even if some participants expressed their interest in creating data collection projects of their own, no additional projects were observed. None of the experiments passed the seeding stage of the democratization model outlined in this thesis. The platform is still available, however, and collecting the instrumented variables to observe if further spread of the usage of the tool happens.

Regarding the usefulness of Flocktracker and its deployment model, the experiments suggest that the usage of digital tools for undertaking community mapping by community organizations can help to give more visibility to the problems for which they advocate solutions. This was observed in the generally positive reactions from the people in the in-person interactions across the experiments, the local media interactions in Solo 3, and the local newspaper attention to the data collection process in Solo 1 and Solo 3.

In addition, the experiments offer other additional useful insights. Regarding the technology, younger and educated people had no problem installing and learning how to use a new Android app on their smart devices; some of the older people (50 years or more), however, who had smartphones did not know how to install new software (apps) and manage them. This reveals the importance of keeping aware of potential gaps in participation in digital community mapping that could skew the generated datasets towards the views and interests of younger participants. The gap can be bridged by creating digital community mapping events with more heterogeneous participants, where more "tech-savvy" individuals can help others setup the software and guide them along the process of digital community mapping. The act of sharing experiences among different user groups could also be helpful in them better understanding their respective points of view.

In terms of possible partnerships, we observed that professors at local universities and their students were interested into testing new tools and approaches to generate urban data. These partnerships might be helpful to democratize data collection. Nonetheless, in the case of digital community mapping, it is important to align incentives so that people participate not only to learn about new tools, but to participate and learn about the digital community mapping subject. The Solo 2 case provided an example of a group of students and professors who were willing to learn about the tools but were not motivated to participate in community data collection. Their attendance at the event may have increased their awareness of digital data collection tools and community mapping, but it was not motivating enough to get them to collect data or create their own projects.

In both settings, local government agencies were apparently open to the idea of using new tools to generate data and to experiment with new approaches to engage with their communities, since they were willing to lend their name and venues. Still, the Solo 1 and Tlalnepantla experiments show that having a digital community mapping event as an addendum to another event organized by the partner government agency is insufficient. Since most people attended with a different agenda in mind (i.e., the main event they were invited to), many of them left immediately after the main event finished and did not to stay for the digital community mapping.

Table 2 provides a summary of the experiments described in this chapter.

Table 2 - Summary of the experiments

	Experiment	Solo 1	Tlalnepantla 1	Solo 2	Solo 3
Project	Key partner	Kota Kita	Planeación y Desarrollo	Kota Kita	Kota Kita
	Project name	Pedalista	mapatlalne	Pedalista	Pedalista
	Number of questions	4	- 12 in a branching project - Only 5 to 6 visible when collecting a survey in field	4	4
Setup	Other partners	- Transportation authority	- Municipal government – Cultura de Paz (Culture of peace) program - Local universities	- Transportation authority - Local universities	- Transportation authority
	Outreach	- In-person coordination with partner's Women on Wheels program. - Invitation to road safety community event just before the Flocktracker event	- In person invitations in local universities. - Outreach to professors at local universities to invite their students - Posters in local universities and in Tlalnepantla neighborhoods - Social media posts on Facebook - Invitation to the safety community event just before the Flocktracker event	- Outreach to local university professors to invite students	- Paid Facebook campaign - Interactions and publications with local newspapers - In-person invitation at Car Free day events
Participation	Incentives for participants	- Learn about digital data collection - Improve transportation infrastructure by learning about it - Collaborating with the Women on Wheels community	- Learn about digital data collection - Improve local safety conditions by learning about them - For students and urban planning professionals, learn more about digital data collection tools	- Learn about digital data collection - Improve transportation infrastructure by learning about it - For urban planning students, learn more about digital data collection tools	- Improve transportation infrastructure by learning about it - Contest to win up to 33 USD in cell phone plan top-ups
	Interaction materials	- Presentation	- Presentation - Leaflet - Poster	- Presentation - Leaflet - Poster	- Leaflet - Poster - Website
	Community event	- Secondary event to Women on Wheels event in Solo	- Secondary event to a safety community event in Tlalnepantla	- Event focused on Pedalista and Flocktracker	- Not held for Pedalista, but outreach as part of Car free day event
	Participants	- ~60 women of 50 or more years old	- 14 people, including: - 4 women 50+ years old - 4 urban planning professionals - 6 students	- 50 city planning undergrad students	- Unknown - Only one person collected data
	Post event participation	- Not observed	- Not Observed	- Not Observed	- 1 upload

	Experiment	Solo 1	Tlalnepantla 1	Solo 2	Solo 3
Data collected	**Pre-survey**	- None, participants left before completing	- Yes, collected with pen and paper surveys	- Yes, collected when installing the app	- Yes, collected when installing the app
	Post-survey	- None	- None	- None	- None
	In-Field data from community mapping	- None observed	- Yes, from initial event	- None observed	- Yes, only one data point from one participant
	Server instrumentation data	- Yes, whole process	- Yes, whole process	- Yes, whole process	- Yes, whole process
Outcomes - Technology	**Positives**		- Young and educated users could setup and use the app with minimal guidance only provided by a leaflet - Older people were able to participate thanks to the help of younger participants	- Young and educated users were able to setup and use the app with minimal guidance only provided by a leaflet	- Mostly unguided usage was possible thanks to the solution of the errors found on the previous iterations
	Negatives	- App was not able to be used by older participants - Large server-side crashes also prevented usage	- Minor layout issues and crashes	- Minor layout issues and crashes.	- Minor layout issues and crashes
Outcomes - Participation process	**Positive**	- Local media attention on community data collection	- Students learned about digital tools for data collection	- Students learned about digital tools for data collection	- Local media attention on community data collection - Widespread social media awareness of community work regarding transportation safety
	Negative	- No data collected - No further participation	- Only a very small area was covered - No further participation	- No data collected - No further participation - Researchers and students cared more about the tool than about participating	- Only 1 upload by a member of the community - No further participation

6 Conclusions

This thesis set out to test a new field data collection tool and understand its potential impacts on the participants in the data collection process and whether that process (and the tool which enabled it) might generate additional new community-based field data collection efforts (i.e., “democratize” data collection). The four field experiments carried out for this thesis research, in two different developing cities in collaboration with two local NGOs, reveal numerous challenges and lessons. This Chapter summarizes the main findings relative to the initial research questions, offers some potential reasons why participation outside the community event was relatively low in all the cases, offers reflections about potential improvements in digital community data collection processes, and ends with some remarks regarding the potential for digital community mapping as a niche to democratize data collection within a community’s sociotechnical landscape.

6.1 Value of Flocktracker and its deployment model toto local community organizations and their communities for gathering in-field data

Flocktracker was able to provide an easy enough tool to create projects for the partner organizations. These organizations were able to setup the projects with very little guidance. However, whether these tools can subsequently be effectively used to motivate community volunteers to participate in data collection remains an open question based on the cases presented here. In other words, tools like Flocktracker could be used by local advocacy organizations, but alone do not guarantee that an appropriate group of community volunteers will be willing to participate, has access to smart phones, and is likely to learn how to use the tool. The cases covered in this research hopefully can help to inform how to make the related processes better.

Flocktracker was easy to understand for younger and educated people but proved to be more challenging to use for older people who are not tech savvy. This apparent “digital divide” needs to be understood better, as most digital data collection projects are likely to face similar challenges. User experience design can alleviate parts of the problem but reaching a digital data collection platform that every person can use requires additional efforts. The experiences in these cases, especially in Tlalnepantla, albeit limited, suggest that co-teaching among participants can help, in this case recruiting younger people to help the older and less tech savvy. This can help to expand inclusiveness.

This research is inconclusive regarding the benefits that digital community mapping with tools like Flocktracker can bring to an organization. The Tlalnepantla case serves as an example of how digital community mapping can be very similar to a more traditional effort in terms of organization and scope; but, similar to the other cases, it did not demonstrate a potential for the tools to help community organizations increase participation, the amount of data collected, or their geographical coverage.

Finally, although the instrumentation of the tool was not utilized quantitatively to answer the research questions in this thesis, it proved to be unobtrusive and it should be useful in other projects with larger communities and more complex sociotechnical landscapes where different organizations and citizens are using the platform for different projects in the same area.

6.1.1 Impact of a digitally enabled community data collection process on participants' perceptions

While pre/post surveys were designed to attempt to measure impacts of the data collection projects on participants' perceptions regarding the role and value of data and data collection, the low survey completion rate and the low participation rate in general rendered the survey information to be of little value. Anecdotally, participants in the projects indicated that they valued more data and evidence as part of the decision-making process of their communities. As discussed in the Tlalnepantla case, participants mentioned that they had a better understanding of the connections between data, understanding, and potential actions. They also mentioned that they would be willing to collaborate more in creating data if the partnerships were clear; for example, if Tlalnepantla's government indicated a willingness to fix all the reported issues that eventually might be revealed (through the data collection) on a map. Only researchers, students and planning professionals that were members of the community expressed a willingness to lead their own projects; in the end, none of them actually did.

6.1.2 Digital community mapping as a means for "democratizing" data collection

Although some participants mentioned a desire to start data collection projects on their own, none did within the timeframe observed for this thesis. The theoretical model for "democratizing" data collection, as presented in Chapter 3 was not validated in this thesis. The four experiments analyzed here have not yet encouraged other projects to start and we can find no conclusive evidence from these projects regarding the design of an initial digital community mapping experience that would encourage others to start their own projects.

Each of the experiments, including their two-week post-event participation periods, only fulfilled the seeding stage of the democratization model envisioned for the Flocktracker platform. Participants' expressed interest in collecting data in other projects in both Tlalnepantla and Solo is encouraging, but no additional data collection has been observed so far. Ongoing monitoring of the platform outside of the scope of this research could reveal evidence of secondary usage spread in Tlalnepantla and/or Solo, however, this might not happen soon, if ever.

6.2 Reflections on the cases and recommendations for future engagements

From the field experiments reported here, aside from the trivial conclusion that creating digital community mapping projects is difficult, one can make various observations about and speculate on reasons for the low levels of participation in terms of number of people attending the events and collecting data.

6.2.1 The Platform

In terms of the Flocktracker software, some evidence emerged that it was too difficult to use, at least for the older adult participants. In the first Solo experiment, this emerged as a major challenge. This challenge was overcome in the Tlalnepantla case, with a smaller group of people and a purposeful design, in which younger people (who had already demonstrated an ability to set up the app and project by themselves) assisted older people in setting up and using the app. Unfortunately, the largest scale outreach effort of the Solo 3 experiment led to less than 30 downloads of the app and evidence of only one user collecting data on her own. We cannot know if these results are due to difficulty in app installation or use or general disinterest in the project among the thousands of potential participants. Regardless, the current state of the app may be more user-friendly, with more explanations about the steps needed to setup the app embedded in the user interface which could also help to ease the process of signing up, collecting data, and understanding the platform in general.

6.2.2 The People

The software challenges relate to demographic and socioeconomic factors, including age, education, and income. Regarding *age*, we might expect younger people to both better understand how to use digital tools, like smartphone apps, and be more aware of participative planning actions in their city or elsewhere through their use of the Internet. On the contrary, as Solo 1 exemplifies, older participants had a more difficult time with the app. Higher *educational attainment* might also play a role, leading, for example, to individuals having a greater understanding about the governance structures in their localities, the NGOs working in their area, and the potential consequences of participation. The Solo 2 and Tlalnepantla experiments provided some supporting evidence in the form of interest expressed by researchers in both cases and the willingness for urban planning professionals to travel more than one hour to participate in Tlalnepantla. Finally, people with higher *income*s may have more free time to learn about their communities and participate when they want to change them. Furthermore, we can safely assume that, even with the rapid penetration of smart devices in developing countries, people with higher incomes are more likely to have expensive smartphones and the know-how to use them. People with higher incomes also likely have higher educational attainment.

6.2.3 The Project/Campaign

In terms of the data collection campaigns themselves, citizens may not have been very interested in the specific subject defined for the digital community mapping projects and/or may not see a clear connection between data collection and urban conditions. Although the data collection projects were focused on specific subjects that members of local NGOs deemed likely to be interesting to local residents, the stated primary purpose of setting up the projects was to move forward the agendas of each NGO: public safety for Planeación y Desarrollo and transportation infrastructure for Kota Kita. This could represent a misalignment between the general public's interests (subjects they might want to participate in) and the agendas that advocacy groups want to move forward. Even if citizens were interested in public safety or transportation infrastructure, they may not have perceived a sense of urgency to work to understand and solve the issue. In other words, people may simply not have felt compelled to spend time on trying to fix the issue now. The solution can wait.

Several factors could contribute to this lack of interest in participation. People may not see a connection between data and change of conditions: even if aware of the need for action, they might not see the potential impact of evidence and data collection on improving understanding and designing solutions. Unfortunately, the after-participation survey, intended to measure this, was not completed in any of the experiments. This is related to the fact that citizens did not participate in the data schema definition; they were only invited to the data collection phase itself. Involving more members of the community in earlier stages of the process might have had motivated some people to participate more. True co-creation in community mapping and data science projects, however, has major implications for the size and capacity of the organizations leading the effort, as well as the time necessary for and the complexity of the overall process (e.g., data schema definition, stakeholder objective alignment, learning about the technology).

More generally, it seems that in the cases in this thesis the incentives were not well aligned. Data collection for the sake of learning about a place might not be a very attractive goal for people that need their problems solved now. The intangible benefit, better data collection which might lead to improvements in the future, were not worth the costs. Even when a tangible incentive was used, the cellphone data plan top-up, this did not seem to resonate with members of the community. This may have been influenced by the token participation of authorities. While all of the cases were backed by government agencies, in manners ranging from them lending their endorsement to incorporating the projects as part of the agenda of a government-led event, the agencies did not participate more deeply nor were they willing to commit to direct actions based on the potential evidence collected. This is understandable, given that the initiatives came from NGOs, but the lack of commitment probably did not positively contribute to people's willingness to participate. Moreover, an additional disincentive may have been people's aversion to data-related projects. Citizens could be against data collection, in principle, and unwilling to share information with third parties and governments. This might be related to privacy concerns (Antón, Julia, & Jessica, 2010). For example, in the experiment in Mexico City, at least one person expressed her reluctance to share her information in the initial survey due to concerns that the government has used this kind of data to retaliate against citizens. Distrust of software providers and governments alike is a perceived negative cost which can hamper the spread of data collection tools.

6.2.4 Potentially beneficial community outreach methods

Paid social media campaigns, as used in the Solo 3 experiment, spread awareness of the digital community mapping projects the fastest, but it was also the least successful in convincing people to participate. In these experiments, holding a community event in a physical place was more effective in attracting people. Nonetheless, although community events might increase awareness of community mapping and digital data collection as a tool to learn about local issues, awareness is not the only condition needed to motivate communities to use the tools for their own needs. The high numbers of social media interactions in Solo 3 relative to the low levels of participation are indicative of this.

For successful community engagement, it is important that the data collection project is part of the initial outreach campaigns of the related community events. In events like those of Solo 1 and Tlalnepantla, it was relatively easy to convince local authorities to include the digital community mapping efforts as part of the agenda of their community events, but people were unlikely to stay for the Flocktracker part, potentially because they attended the advertised main event and did not expect to have to stay longer. Obviously, timely advanced notice for potential participants also is key for higher attendance levels. Both Tlalnepantla and the Solo 2 experiments involved specific outreach for the data collection event, but the former was announced with very short notice (less than a week). Some people in the Tlalnepantla case communicated and told us they were not able to come given the short notice; Solo 2, with a more organized outreach, had 50 attendants.

6.2.5 Potentially impactful partnerships within the community

In general, the NGOs and other organizations that participated in the experiments expressed their interest in using new technologies and approaches for generating new datasets. This initial interest can be beneficial in spreading the word about new projects in an area. This could be suggestive of an institutional environment supportive of spreading the use of digital data collection tools. Despite their positive interest, they also faced funding problems and tech savviness challenges.

Of all the members of the community that responded to the outreach efforts, researchers and students seemed more likely to be interested in learning about community data collection. Thus, partnering to organize digital community mapping with universities and research institutions could be positive if the local organization already has good ties with them. At the same time, it can also bias the data collection process towards focusing more on topics of interest to the researchers and students rather to the interests of the rest of the community or the organizers. In the process of democratizing data collection, universities can be good partners to start the spreading of tools like Flocktracker. With enough time, students could take the tools and use them for objectives other than education-led research, such as activism, citizen science, other community work, and their own professional output.

Although having the endorsement of local governments gave the projects more legitimacy and the organized events were able to take advantage of the logistics and resources of their community outreach meetings, some people in Tlalnepantla 1 were so dissatisfied with their local government that they did not want to participate. As mentioned, one person even expressed fear about government retaliation if she shared negative views or personal data. Another problem mentioned was the lack of explicit willingness by the governments to respond to people's reports and improve local conditions. For the government partners in Tlalnepantla and Solo, the events were treated mostly as interesting pilots. For creating a space more conducive to participation, government partners should lend not only their names or offices, but their commitment to use the data collection to inform their actions and respond to the identified community needs.

6.3 Moving forward

This section summarizes some key lessons for future consideration in developing better digital community mapping events.

6.3.1 Partnerships

For government partnerships to be successful, the local government partner must be willing to participate more than in name or in simply organizing the community event. There should be a direct and clear expression of willingness to listen to people's concerns from both the digital community mapping event and the content of the data collected.

For research and academic partners, further use of the tool or techniques for their research needs to take a secondary position behind the objective of using the data collection project to help achieve local community and NGO goals. If students are to be invited to the event, they should be invited as participants that can collaborate and create a digital map with their communities rather than just inviting them to learn about the new technologies. Learning about the new technology should happen by using it.

6.3.2 Event

The community events proved to be useful to share the motivations of the digital community mapping process. When matched with hands-on data collection by the community, at least an initial understanding of the technology and some shared experiences can be built. During such an event, introduction of the general reasons for the project should be short, to allow time for hands-on data collection on the day of the event itself. A follow-up session for sharing findings and learnings after the process, like the one done in Tlalnepantla, can be very helpful to understand better what people learned, to encourage them to continue collecting data, and to study the process to iterate with better events later.

Bringing together a mix of people (e.g., young and older) can be a way to bridge the digital divide that may prevent people from some groups from participating. By helping each other and working closely, people from different backgrounds or demographics (e.g., age groups) can also share their different views and experiences.

6.3.3 Outreach

The event's time and place should be set as early as possible. Prior notice of two weeks or more would be advisable so that community members can adequately prepare to attend. Social media outreach is a good tool for spreading the word about the event, but it needs to have clear calls for action to participate. While not tried in any of the experiment cases in this thesis, a combination of paid social media outreach and a community event with hands-on data collection seems like a promising possible approach. The same social media platforms can be used to remind people to participate after the initial event.

6.3.4 Incentives

The incentives that were considered to encourage people to keep participating after the initial event did not work. For the contest that was set up for Solo 3, even if the award was very clear and upfront, it did not generate much response. Perhaps the value was too low or people simply were not interested in the data plan top-ups that were awarded. Better involvement of the government partners could be useful to motivate people. This would likely entail the government committing to move forward in solving the issues raised by the data collection, in collaboration with the NGO and the community. Better understanding of the community is needed before setting up a contest with prizes since the prizes might or might not (as in the case with Solo 3) motivate participation.

6.3.5 Data collection seed project

The subject of the data collection process itself needs to be well aligned with the interests of the NGO and their partners, but also with an understanding of the community's interests. In the cases studied here, the community was working with projects that were set-up by the NGOs and me. In this case, even if the data collected was a participatory process, the project that defined the structure of the data was not. To have community members more involved in co-creating the data collection project design itself might be a good way to generate more people interested in the community data collection process.

6.3.6 Post event participation

As mentioned earlier, social media outreach can be useful to remind people to continue participating. Another motivator could be setting up a second, post-data collection, event and/or activities, where the full dataset is analyzed by members of the community and the local organizational partners to share learnings and strategies to solve the issues identified during the process. Social media can also be used to remind people who already participated in the initial event to keep doing so on their own time. With more development, reminders to participate could be embedded in Flocktracker or similar apps.

6.4 Digital community mapping as a niche to democratize data collection

This process made it clear that data savviness is limited at the organizational, governmental and citizen levels in Solo and Mexico City. Both NGOs and governments in Tlalnepantla and Solo are taking steps to increase data-driven approaches for their strategies and their organizational capacities to design and implement data collection and analysis technologies. Still, citizens have not been very involved in this yet and some population groups, like at least some older adults, have a high barrier to entry. Digital community mapping holds promise in helping people understand why data is becoming more and more important in the governance process of their communities. This realization could be the precursor of more data driven community-led projects in Tlalnepantla and Solo.

Although the digital community mapping experiences in this thesis were not very successful in scale and depth, the iterative process of the experiments did prove to be conducive to improving the approaches undertaken when designing and deploying the projects. Subsequent projects can build from the lessons of the four experiments researched here. Ultimately, embedding digital data collection into the sociotechnical landscape of a community will require continuous learning about and improvement of both the technologies involved and the strategies used to share the knowledge about the technologies' use.

Bibliography

Lütfi Süzen, M., & Doyuran, V. (2004). Data driven bivariate landslide susceptibility assessment using. *Engineering Geology, 71*, 303-321.

Alawadhi, S., & Scholl, H. (2013). Aspirations and realizations: the smart city of Seattle. *46th Hawaii International Conference on System Sciences, IEEE Proceedings*, (pp. 1695-1703). Wailea, Maui, HI.

Amsden, J., & VanWynsberghe, R. (2005). Community mapping as a research tool with youth. *Action Research, 3*(4).

Antón, A. I., Julia, E. B., & Jessica, Y. D. (2010). How Internet Users' Privacy Concerns Have Evolved Since 2002. *IEEE Security & Privacy, 8*(1), 21-27.

Arnstein, S. R. (1969, July). A ladder of Citizen Participation. *Journal of the American Planning Association, 34*(4), 216-224.

Baak, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society, 1*(11).

Badan Perencanaan Pembangunan Nasional. (2009). *Pencapaian Sebuah Perubahan - Evaluasi 4 Tahun Pelaksanaan Rpjmn 2004-2009.* Jakarta, Indonesia: Badan Perencanaan Pembangunan Nasional.

Badan Pusat Statistik (Statistics Indonesia). (2018, 12 19). *Indeks Pembangunan Manusia menurut Provinsi, 2010-2017 (Metode Baru)*. Retrieved from Badan Pusat Statistik: https://www.bps.go.id/linkTableDinamis/view/id/1211

Badan Pusat Statistik (Statistics Indonesia). (2018, 12 12). *Penduduk Indonesia menurut Provinsi 1971, 1980, 1990, 1995, 2000 dan 2010*. Retrieved from Badan Pusat Statistik: https://web.archive.org/web/20130701144756/http://www.bps.go.id/tab_sub/view.php?kat=1&tabel=1&daftar=1&id_subyek=12&notab=1

Batty, M. (2008). The Size, Scale, and Shape of Cities. *Science, 319*(5864), 769-71.

Batty, M., Axhausen, K., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., . . . Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics, 214*, 481-518.

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

Butts, K. (2014). *Iterative Methods in Adapting Mobile Technologies for Data Acquisition, A Case Study in St. Louis, Missouri (Master's Thesis).* Cambridge, Massachusetts: Massachusetts Institute of Technology.

Caniëls, M. C., & Romijn, H. A. (2008). Strategic niche management: towards a policy tool for sustainable development. *Technology Analysis and Strategic Management, 20*(2), 245-266.

Carvalho, L. (2014). Smart cities from scratch? a socio-technical perspective. *Cambridge Journal of Regions, Economy and Society*.

Chakraborty, A., Wilson, B., Sarraf, S., & Arnab, J. (2015). Open data for informal settlements: Toward a user' s guide for urban managers and planners. *Journal of Urban Management, 4*, 74-91.

Chevalier, J. M., & Buckles, D. J. (2013). *Participatory Action Research - Theory and Methods for Engaged Inquiry.* Routledge.

Ching, A., Zegras, C., Kennedy, S., & Mamun, M. (2012). A User-Flocksourced Bus Experiment in Dhaka: New Data Collection Technique with Smartphones. *Transportation Research Board Annual Meeting.*

Cosio Zavala, M. E. (1995). *Changements de Fécondité au Mexique el Politiques.* Paris, France: Editions L'Harmattan.

Crooks, A., Pfoser, D., Jenkins, A., & Croitoru, A. (2015). Crowdsourcing urban form and function. *International Journal of Geographical Information Science, 29*(5), 1-22.

Delgado, J., Larralde, A., & Anzaldo, C. (1999). La Corona Regional de la Ciudad de México. Primer Anillo Exterior en Formación. In J. Delgado, & B. Ramírez, *Transiciones. La Nueva Formación Territorial de la Ciudad de México. México: Programa de Investigación Metropolitana-UAM* (pp. 171-194). Plaza y Valdés.

Donovan, K. (2012). Seeing Like a Slum: Towards Open, Deliberative Development. *Georgetown Journal of International Affairs, 13*(1), 97-104.

Ferreira, N., Lage, M., Doraiswamy, H., Vo, H., Wilson, L., Werner, H., . . . Silva, C. (2015). Urbane: A 3D framework to support data driven decision making in urban development. *2015 IEEE Conference on Visual Analytics Science and Technology (VAST).* Chicago, IL, USA: IEEE.

Flocktracker. (2018, 10 18). *Use cases*. Retrieved from Flocktracker.org: https://www.flocktracker.org/use-cases-2/

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*(31), 1257-1274.

Goodchild, M. F. (2011). *Citizens as Sensors: The World of Volunteered Geography.* John Wiley & Sons.

Goodspeed, R. (2014). Smart cities: moving beyond urban cybernetics to tackle wicked problems. *Cambridge Journal of Regions, Economy and Society*.

Google. (2018, 05 24). *Material Design for Android*. Retrieved from Android Developers: https://developer.android.com/guide/topics/ui/look-and-feel/

Gramlich, E. M. (1994). Infrastructure Investment: A Review Essay. *Journal of Economic Literature, 32*(3), 1176-1196.

Greenfield, A. (2012, March 6). *Week 61: Spontaneous order (and value) from the bottom up*. Retrieved from Urban Scale: http://urbanscale.org/news/2012/03/06/week-61-spontaneous-order-and-value-from-the-bottom-up/

Harley, J. B. (1989). Deconstructing the map. *Cartographica: The international journal for geographic information and geovisualization, 26*(2), 1-20.

Hollands, R. G. (2014). Critical interventions into the corporate smart city. *Cambridge Journal of Regions, Economy and Society*.

Hwang, Y. H., & Roscoe, C. J. (2017). Preference for site conservation in relation to on-site biodiversity and perceived site attributes: An on-site survey of unmanaged urban greenery in a tropical city. *Urban Forestry & Urban Greening, 28*, 12-20.

International Telecomunications Union. (2018, 12 12). *Statistics*. Retrieved from International Telecomunications Union: https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly, 28*(4), 446-456.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management, 29*(4), 258-268.

Kemp, R., Schot, J., & Remco, H. (1998). Regime shifts to sustainability through processes of niche formation: The approach of strategic niche management. *Technology Analysis & Strategic Management, 10*(2), 175-198.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal, 79*, 1-14.

Klosterman, R. E. (1994). Large-Scale Urban Models Retrospect and. *Journal of the American Planning Association, 60*(1), 3-6.

Kota Kita. (2018, 12 12). *Kota Kita*. Retrieved from Projects: http://www.kotakita.org/projects

Lawson, A. B., Biggeri, A., Böhning, D., Lesaffre, E., Jean-Fran, V., & Bertollini, R. (1999). *Disease Mapping and Risk Assessment for Public Health.* New York, NY, USA: 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.

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

Mandinach, E. B., Honey, M., & Light, D. (2006). A Theoretical Framework for Data-Driven Decision Making. *Annual meeting of the American Educational Research Associatio.* San Francisco, CA, USA.

Millar, E. E., Hazell, E. C., & Melles, S. J. (2018). The 'cottage effect' in citizen science? Spatial bias in aquatic monitoring programs. *International Journal of Geographical Information Science*.

Perkins, C., & Thomson, A. Z. (n.d.). COMMUNITY MAPPING: CHANGING LIFESTYLES THROUGH PARTICIPATION.

Pinder, D. (1996). Subverting Cartography: The Situationists and Maps of the City. *Environment and Planning A: Economy and Space, 28*(3).

Planeación y Desarrolo S.C. (2018, 12 12). *Proyectos*. Retrieved from Planeación y Desarrollow: https://web.archive.org/web/20161201231705/http://www.planeaciondesarrollo.com.mx/proyectos/

Poushter, J. (2016). *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies But advanced economies still have higher rates of technology use.* Washington, DC: Pew Research Center.

Programa de las Naciones Unidas para el Desarrollo en México. (2015). *Índice de Desarrollo Humano para las entidades federativas, México 2015.* Programa de las Naciones Unidas para el Desarrollo en México.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data, 1*(1).

Rambaldi, G., Kwaku Kyem, P. A., Keith McCall, M., & Weiner, D. (2006). Participatory Spatial Information Management and Communication in Developing Countries. *Electronic Journal of Information Systems in Developing Countries, 25*(1).

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*(4), 155-169.

Savas, E. (1970). Cybernetics in City Hall. *Science*(168), 1066.

Schot, J., & Geels, F. W. (2008). Strategic niche management and sustainable innovation journeys: theory, findings, research agenda, and policy. *Technology Analysis & Strategic Management, 20*(5), 537-554.

The World Bank. (2018, 12 2). *Urban population (% of total)*. Retrieved from The World Bank - Data: https://data.worldbank.org/indicator/sp.urb.totl.in.zs

Truffer, B., Metzner, A., & Hoogma, R. (2002). The Coupling of Viewing and Doing: Strategic Niche Management and the electrification of individual transport. *Greener Management International - "Foresighting and Innovative Approaches to Sustainable Development Planning"*.

U.S. Department of Health & Human Services. (n.d.). *User-Centered Design Basics*. Retrieved 08 21, 2018, from usability.org: https://www.usability.gov/what-and-why/user-centered-design.html

United Nations. (2018, 12 2). *World Population Prospects 2017*. Retrieved from United Nations - Department of Economic and Social Affairs: https://population.un.org/wpp/

Werlin, H. (1999). The Slum Upgrading Myth. *Urban Studies, 36*(9), 1523-1534.

White, R., & Engelen, G. (2000). High-resolution Integrated Modelling of Spatial Dynamics of Urban and Regional Systems. *Computers Environment and Urban Systems, 24*(5), 383-400.

Zavala de Cosio, M. E. (2014). La transición demográfica en Mexico (1895-2010). In C. Rabell, *Los mexicanos Un balance del cambio demográfico* (pp. 80-114). Fondo de Cultura Economica.

Zegras, P. C., Butts, K., Cadena, A., & Palencia, D. (2015). Spatiotemporal dynamics in public transport personal security perceptions: digital evidence from Mexico City's periphery. In N.

Thomopoulos, M. Givoni, & P. Rietveld, *ICT for Transport: Opportunities and Threats* (pp. 115-139).

Zegras, P. C., Eros, E., Butts, K., Resor, E., Kennedy, S., Ching, A., & Mamun, M. (2015). Tracing a path to knowledge? Indicative user impacts of introducing a public transport map in Dhaka, Bangladesh. *Cambridge Journal of Regions, Economy and Society, 8*(1), 113-129.

Zetter, R., & De Souza, F. A. (2010). Understanding Processes of Informal Housing: Appropriate Methodological Tools for a Sensitive Research Area. *International Planning Studies, 5*(2), 149-164.

Annex 1 – Survey instruments

Initial survey

Flocktracker: Collaborative data collection

Pilot App Phase using Smartphones

Questions:

1. What is your email? __________________
2. What is your Flocktracker username? (the one you just created in the app) ____________________
3. What is your gender:
    a. Female
    b. Male
    c. Prefer not to answer
4. What is your age?
    a. <18
    b. 18-25
    c. 25-35
    d. 35-45
    e. Older than 45
5. What is your educational attainment?
    a. No school
    b. Primary School
    c. Secondary School
    d. High School
    e. Technical school
    f. College
    g. Graduate Studies

6. Did you know about Flocktracker before coming to this event?
    1. I have never heard of it
    2. I have heard of it, but never used it
    3. I have used it occasionally
    4. I have used it several times
    5. I have used it many times.
7. Do you know how to play Pokémon Go?
    1. I have never heard of it
    2. I have heard of it, but never played it
    3. I have played it occasionally
    4. I have played it often
    5. I have played it many times

The following questions are about your familiarity with computer software:

	I have never heard of it	I have heard of it, but never used it	I use it monthly	I use it weekly	I use it daily
8. Do you know how to use Microsoft Excel?					
9. Do you know how to use ArcGIS?					
10. Do you know how to use Twitter?					
11. Do you know how to use Facebook?					
12. Do you know how to use Google Maps?					

Please read the following statements and indicate how much you agree or disagree with them:

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
13. Citizens have an important role to play in generating data and information to be used for decision-making					
14. The government should increase the availability of data for citizens					
15. Data can play an important role in helping citizens better understand their communities					

16. Choose the 5 types of data would you prefer to be more available for your town (city)?

- ☐ Business data
- ☐ Crime and justice data
- ☐ Environment and weather data
- ☐ Housing data
- ☐ Sanitation infrastructure data
- ☐ Energy infrastructure data
- ☐ Transportation infrastructure data (e.g., roads, paths)
- ☐ Transportation services data (e.g., buses, taxis)
- ☐ Government income and spending data
- ☐ None of the above
- ☐ Other __________

17. Choose the 3 main challenges to better use of data in local governance
    - ☐ Insufficient capacity, skills
    - ☐ Lack of appropriate data
    - ☐ Lack of data collection methods
    - ☐ Inability to use data
    - ☐ Poor data and information sharing
    - ☐ Limited possibilities for citizens to participate in data collection
    - ☐ Low recognition of potential value of crowdsourcing data
    - ☐ Weak legal and regulatory framework.
    - ☐ Other ________

18. Choose the three (3) most important things you believe are most important for your local government to make better decisions?
    - ☐ Increase openness and transparency
    - ☐ Facilitate public participation in decision making
    - ☐ Facilitate public participation in data collection
    - ☐ Deliver public services more effectively with better processes and information
    - ☐ Enable citizens to take more informed personal decisions
    - ☐ Gather better data and information
    - ☐ Other ______________

Thank you!

Final survey

Please read the following statements and indicate how much you agree or disagree with them:

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
1. Citizens have an important role to play in generating data and information to be used for decision-making					
2. The government should increase the availability of data for citizens					
3. Data can play an important role in helping citizens better understand their communities					

4. Choose the 5 types of data would you prefer to be more available for your town (city)?

- ☐ Business data
- ☐ Crime and justice data
- ☐ Environment and weather data
- ☐ Housing data
- ☐ Sanitation infrastructure data
- ☐ Energy infrastructure data
- ☐ Transportation infrastructure data (e.g., roads, paths)
- ☐ Transportation services data (e.g., buses, taxis)
- ☐ Government income and spending data
- ☐ None of the above
- ☐ Other __________

5. Choose the 3 main challenges to better use of data in local governance
   - ☐ Insufficient capacity, skills
   - ☐ Lack of appropriate data
   - ☐ Lack of data collection methods
   - ☐ Inability to use data
   - ☐ Poor data and information sharing
   - ☐ Limited possibilities for citizens to participate in data collection
   - ☐ Low recognition of potential value of crowdsourcing data
   - ☐ Weak legal and regulatory framework.
   - ☐ Other ________

6. Choose the three (3) most important things you believe are most important for your local government to make better decisions?
   - ☐ Increase openness and transparency
   - ☐ Facilitate public participation in decision making
   - ☐ Facilitate public participation in data collection
   - ☐ Deliver public services more effectively with better processes and information
   - ☐ Enable citizens to take more informed personal decisions
   - ☐ Gather better data and information
   - ☐ Other ______________

Annex 2 – List of variables for server-side instrumentation for usage tracking

Server Level

Variable name	Description
server_total_authenticated_calls	S1. Number of calls to the server - How many times the user used any endpoint authenticated in the server (Successful + failed calls)
server_total_authenticated_successful_calls	S2. Number of successful calls to the server
server_total_authenticated_failed_calls	S3. Number of failed calls to the server
server_total_authenticated_server_error_failed_calls	S4. Number of client based failed calls to server
server_total_authenticated_client_error_failed_calls	S5. Number of server based failed calls to server
server_last_signup_time	S6. Last time a user joined the server - Recent most sign up time
server_first_signup_time	S7. First time a user joined the server - First time sign up
server_first_project_creation_time	S8. First time a project was created
server_last_project_creation_time	S9. Last time a project was created
server_first_survey_creation_time	S10. Time of first survey creation
server_last_survey_creation_time	S11. Time of last survey creation
server_first_trip_creation_time	S12. Time of first trip creation
server_last_trip_creation_time	S13. Time of last trip creation
server_first_counter_session_creation_time	S14. Time of first counter session creation
server_last_counter_session_creation_time	S15. Time of last counter session creation
server_total_surveys	S16. Total number of surveys (not including deletion)
server_total_trips	S17. Total number of tripIDS (not including deletion)
server_total_counter_sessions	S18. Total number of counter sessions (not including deletion)
server_total_existing_users	S19. Total number of users (including deletion)
server_total_registered_users	S20. Total number of users overall (not including deletion)
server_current_total_surveys	S21. Total number of surveys (including deletion)
server_current_total_trips	S22. Total number of trips IDs (including deletion)
server_current_total_counter_sessions	S23. Total number of counter sessions (including deletion)
server_total_projects	S24.Total number of projects (not including deletion)
server_current_total_projects	S25. Total number of projects (including deletion).

User Level

Variable name	Description
user_total_calls_to_server	U1. Number of calls to the server - How many times the user used any endpoint authenticated in the server (Successful + failed calls)
user_total_successful_calls_to_server	U2. Number of successful calls to the server
user_total_failed_calls_to_server	U3. Number of failed calls to the server
user_total_client_error_calls_to_server	U4. Number of client based failed calls to server
user_total_server_error_calls_to_server	U5. Number of server based failed calls to server
user_signup_time	U6.Time of joining the platform - Sign up time
user_first_project_creation_time	U7. First time of project creation - First time "owner" status - First time the user uploads a project by herself
user_last_project_creation_time	U8. Time of last project creation
user_first_project_participation_time	U9. Time of first being part of a project- First time the user is part of a project in any way (incl OWNER, joining or being added)
user_first_project_joining_time	U10. Time of first project joining - First time the user joins a project (Using the join proj endpoint)
user_last_project_joining_time	U11. Time of last project joining
user_first_project_being_added_to_time	U12. Time of first being added by someone to a project
user_last_project_being_added_to_time	U13. Time of last being added by someone to a project
user_first_survey_creation_time	U14. Time of first survey creation
user_last_survey_creation_time	U15. Time of last survey creation
user_first_trip_creation_time	U16. Time of first trip creation
user_last_trip_creation_time	U17. Time of last trip creation
user_first_counter_session_creation_time	U18. Time of first counter session creation
user_last_counter_session_creation_time	U19. Time of last counter session creation
user_first_adding_member_to_project_time	U20. Time of first adding someone else to a project (as the user being manager or owner)
user_last_adding_member_to_project_time	U21. Time of last adding someone else to a project (as the user being manager or owner)
user_total_projects_participating	U22. Number of projects the user is part of (Counting deletions as -1) - Currently (depends on project creation and deletion) • User joins a project • User/Owner/manager adds the user to the project • User/manager removes from the project • User creates a project • Project is deleted

Variable name	Description
user_total_projects_participation	U23. The sum of all the projects, not counting deletions
user_total_surveys	U24. Number of total surveys (not including deletion)
user_total_tracking_points	U25. Number of total tracking points (not including deletion)
user_total_trips	U26. Number of trips (not including deletion)
user_total_counters	U27. Number of total counters (not including deletion)
user_total_counter_session	U28. Number of counting sessions (not including deletion)
user_total_projects_participation	U29. Total number of projects the user has been part of (not including deletion) • By creating • By being added • By adding him/herself
user_total_projects_added_to	U30. Total number of projects the user was added to (not including deletion)
user_total_projects_joined	U31. Total number of projects the user joined (not including deletion) By pressing the join button
user_total_projects_created	U32. Total number of projects created (not including deletion)
user_total_members_added	U33. Number of times the user added someone else to a project
user_total_current_surveys	U34. Number of total surveys (including deletion)
user_total_current_tracking_points	U35. Number of total tracking points (including deletion)
user_total_current_trips	U36. Number of trips (including deletion)
user_total_current_counters	U37. Number of total counters (including deletion)
user_total_current_counter_sessions	U38. Number of total counting sesssions (including deletion)
user_total_current_projects_added_to	U39. Total number of projects the user was added to (including deletion)
user_total_current_projects_joined	U40. Total number of projects the user joined (including deletion) By pressing the join button
user_total_current_projects_created	U41. Total number of projects created (including deletion)

Project Level

Variable name	Description
project_participation	P1. Number of people in the project (taking deletions into account)
project_total_current_survey	P2. Number of total surveys (taking deletions into account)
project_total_current_tracker_points	P3. Number of total tracking points (including anything that has a breadcrumb element in a trip) (taking deletions into account)
project_total_current_trips	P4. Number of total trips (taking deletions into account)
project_total_current_counter_points	P5. Number of total counters (taking deletions into account)
project_total_current_counter_sessions	P6. Number of total counter sessions (taking deletions into account)
project_creation_time	P7. Time of project creation
project_last_edit_time	P8. Last time of project edition (using project builder)
project_first_survey_upload_time	P9. First time of survey creation
project_last_survey_upload_time	P10. Last time of survey creation
project_first_trip_upload_time	P11. Time of first trip creation
project_last_trip_upload_time	P12. Time of last trip creation
project_first_counter_upload_time	P13. Time of first counter session creation
project_last_counter_upload_time	P14. Time of last counter session creation
project_total_survey	P15.Number of total surveys (not taking deletions into account)
project_total_tracking_points	P16.Number of total tracking points (not taking deletions into account)
project_total_trips	P17.Number of total trips(not taking deletions into account)
project_total_counters	P18.Number of total counters (not taking deletions into account)
project_total_counter_session	P19.Number of total counter sessions (not taking deletions into account)