Mapjack: A Mobile based Wiki for Collaborative Map Making

by

Jude Ntabathia

BSc BIT, Strathmore University

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2015

Author..... Signature redacted ....
Program in Media Arts and Sciences
May 8, 2015

Certified by........ Signature redacted ....
Ethan Zuckerman
Principal Research Scientist
MIT Media Lab
Thesis Supervisor

Accepted by...... Signature redacted ....
Pattie Maes
Academic Head, Program in Media Arts and Sciences

Mapjack: A Mobile based Wiki for Collaborative Map Making

by

Jude Ntabathia

BSc BIT, Strathmore University

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning,

in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Abstract

The internet has enabled people with few or no formal qualifications to create, share and disseminate cartographic information. One of the best platforms that has empowered citizens to create cartographic data is OpenStreetMap. It has been ten years since OpenStreetMap came into existence and despite generating a large corpus of user contributed cartographic data, significant gaps still exist between countries that have poor access to internet and computers and nations that are highly connected digitally. Mapjack is a mobile application that seeks to bring about more contributions by people who have largely been disenfranchised within this realm of volunteered geographic information. Disenfranchised not because they are not willing to participate, but because of the lack of access to tools and devices to participate. Mapjack enables communities to use mobile phones to create and modify their own spatial identities. Spatial identities refer to how communities want to be spatially represented within the digital realm. These spatial identities are maps showing what communities consider as important and fundamental in representing their geographical space therefore forming a spatial digital identity. A team of two communities that are frequent contributors to OpenStreetMap have participated in evaluating the tool. This thesis highlights that the use of offline map rendering and spatial context play a very important role in consumption and contribution to OpenStreetMap.

Thesis Supervisor: Ethan Zuckerman

Title: Principal Research Scientist, Center for Civic Media

MIT Media Lab

Mapjack: A Mobile based Wiki for Collaborative Map Making

by

Jude Ntabathia

Signature redacted

Thesis Advisor...

Ethan Zuckerman
Director, Center for Civic Media
Principal Research Scientist
Massachusetts Institute of Technology

Signature redacted

Thesis Reader...

Chris Schmandt
Director, Living Mobile
Principal Research Scientist
Massachusetts Institute of Technology

Signature redacted

Thesis Reader...

Sarah Williams
Assistant Professor of Urban Studies and Planning
Director, Civic Data Design Lab
Massachusetts Institute of Technology

Acknowledgements

I would like to thank my thesis supervisor Ethan Zuckerman for his guidance and support over the past two years and for providing me with the opportunity to be here at the lab. Thanks to my colleagues at the Centre for Civic Media specifically Gorgon Mangum, Erhardt Graeff and Lilia Kilburn for their work in reading through this thesis. I would also like to thank Nate Matias for his advice around measuring and evaluating the success of this tool. A special mention goes out to members of Wildlife Works, MapMathare, Spatial Collective and Swahili Box for their invaluable support in deploying the tool with their communities, especially community leaders Ahmed Maawy of SwahiliBox and Simon Kokoyo of MapMathare. Lastly to many members both active and passive of the OpenStreetMap community.

Table of Contents

1 Introduction

1.1 Overview

Mapjack is a mobile based application that seeks to bring about more inclusivity within the realm of volunteered geographic information. In particular it seeks to empower citizens in emerging economies with the ability to contribute to OpenStreetMap. It is intended to be used by citizens whose primary device is an Android smartphone. According to the GSMA, devices that ran Android are estimated to take more than 70 percent of the smartphone market share. This number is still increasing with the latest statistic being 950.5 million devices in the year 2014(GSMA). The application enables citizens to create, share and disseminate vector based cartographic information with people around them and with the general public. Simultaneously, it provides users with perspectives on spatial representation and helps them mould their spatial identities. Spatial identity is what I consider how communities identify themselves based on place constructs.

These forms of spatial identities and representation have a very important role in helping us understand how communities grow and the current state of such communities. These records of cartographic data show how such communities have changed over time. Volunteered Geographic Information, or VGI in short, helps us understand social, physical and economic facets of the communities we are involved in. It helps us understand these communities from different facets, facets which have political, social as well as economic implications.

One of the platforms which best represents this concept of VGI is OpenStreetMap. OpenStreetMap has liberalized access to cartography, which before was a preserve of state agencies, and brought about the rise of the citizen cartographer.

It has been ten years since OpenStreetMap came into existence. Started by Steve Coast in 2004, OpenStreetMap (OSM) was inspired by the success of Wikipedia. The restrictive nature then of proprietary cartographic data and the lack of freely available data were the challenges the project sought to address by creating an open editable map of the world. Ten years down the line, OSM is a global movement of mapping experts, hobbyists and civic actors who are using OSM's main infrastructure to share and collect spatial data. OSM to date has slightly over two million contributors contributing spatial data through different forms. This data is largely collected through manual surveys, tracing using satellite imagery or through the use of GPS devices.

Though OSM has been a phenomenal tool in empowering people with very little cartographic knowledge in creating cartographic data, certain challenges remain and continue to affect how citizen mappers participate. One of the large challenges facing OSM as an ecosystem is the lack of tools available to empower people with mobile phones to participate. Groups for whom the mobile phone is the primary device for accessing the internet have been disadvantaged in contributing to this key database of cartographic data. Moreover, this lack of access to data and the tools to contribute is compounded by the native architecture and mechanisms of OSM. Ten years ago when the project was started, the software was architected to be used over the web via always on connectivity and desktop computers. These architectural decisions affect how data was structured and how these processes assumed the need for massive computing power given the resource intensive nature of cartographic content. Given that computing was centralized, spatial representation was also centralized, leaving no room for communities to define their spatial identities.

During my work with local communities prior and during the course of this thesis, I came to the realization that we as professional cartographers make certain assumptions around matters of spatial representation. Assumptions such as the symbology of what should be spatially represented and not or assumptions on what should be given importance when representing objects on a map. More importantly, we as cartographers fail to involve users in the design of such maps and hence have a tendency to create culturally insensitive spatial representations. To provide an example from my fieldwork: a local community based in Embu, Kenya was interested in adding their shrines to OpenStreetMap. Their main motivation behind adding these shrines was as a mechanism to highlight tourism information and to provide a dataset of cultural history. During the process of adding the shrines to OSM, many members of this community were displeased with how the shrines were represented on OSM. The shrines were represented using the default OSM symbology which equated such amenities to already existing religious structures such as churches or mosques.

To the members of the community in Embu,this was a misrepresentation of their spatial identity. The main reason as to why such shrines took up default symbologies is largely due to the analogous nature of OpenStreetMap where it assumes the model of one schema fits all:

since symbols used to represent churches and mosques are adequate for representing religious sites in Europe or North America, they assumed the same symbols could be used to represent similar sites in Kenya. Efforts to work with already existing OSM data working groups did not bare fruit since moderators felt the need to keep a uniform structure of the global world map dataset. One of the options suggested was to add the data and provide points of interest with unique metadata tags. This way we would be contributing data but sadly most could not interact with the data on the default OpenStreetMap basemap. For instance this would entail adding the tag "traditional-shrine" tag to all the points we added. On reviewing the map, most of these points would not be rendered since they are not part of the global representation and symbology schema. If we are to empower communities to mould their spatial identities, this has to change. Making these changes to the OSM project is a primary motivation behind Mapjack, which seeks to make participatory mapping and OSM in general more inclusive.

1.2 Goals for Building Mapjack

The goal of this thesis project is to build an offline mapping and data collection tool for citizens and enthusiasts who are current OSM users and for whom the mobile phone is the primary device at their disposal. With the Mapjack tool, this group of users will be empowered to edit OpenStreetMap more effectively. Offline mapping helps users in environments where they have little or no access to the internet, providing them with the same mapping environment independent of bandwidth needs. There are some mobile applications which can be used in editing OpenStreetMap data, though all are internet based and rely on persistent internet connections. This dependence on persistent internet connections means that tools like Vespucci, OSMTracker or OpenDataKit are difficult to operate in offline environments or fail to provide spatial context for users when making edits. In markets where internet is not easily available and expensive, this is a barrier to participation by those who would like to contribute using their mobile phones.

The overall goal of Mapjack is bringing about more inclusivity to OpenStreetMap by providing a tool that makes it easier to contribute to OSM through the use of smart phones. It seeks to empower traditionally marginalized groups in the map-making process and seeks to change how those who are marginalized can weave their spatial identities. Reasons for this marginalization include lack of access to the internet, the cost to access internet and poorly designed tools to foster consumption and contributions to OSM.

1.3 Overview of thesis

This thesis documents and highlights some background work and related literature around citizen mapping, its origins, motivations and some of the theoretical studies done around VGI. It then highlights some of the challenges facing OpenStreetMap and studies conducted to make the case for the need of such a tool. This is followed by an assessment of existing tools and some of the design principles behind building Mapjack. The development plan and decisions behind the development are highlighted and experiences from the deployments carried out. It then finishes by looking at challenges faced, an evaluation of the success of this thesis and possible future work.

2 Literature Review

Public Participatory Geographic Information System or PPGIS in short is a term that came into the fore in 1996 during the National Center for Geographic Information and Analysis meeting where it was described as the field of geographic information systems that seeks to enhance public participation and foster the empowerment of non-governmental organisations, grassroots groups and local communities (Sheppard et al. 1999, Sieber 2006). Inherent in the idea of PPGIS is the concept of changing power dynamics in the presentation and analysis of data to help communities make informed decisions about civic issues(Sieber 2007).

Volunteered geographic information (VGI) is a term coined by Michael Goodchild to refer to the widespread engagement of citizens with very little formal qualifications in the creation of geographic information (Goodchild, 2007). The creation of such information has largely been a preserve of government institutions, nonprofits and subject matter experts. VGI as a practice has been employed in different fields simply because of its participatory nature, spanning areas such as urban planning to election monitoring. The ability citizens currently have to create, share and disseminate spatially referenced data not only empowers local communities but also helps shape decisions that such communities could take. Hence, such user generated geographic information has implications for social, economic and political processes.

What is the difference between VGI and PPGIS? In certain cases, both terms have been used synonymously. Some scholars such as Tulloch argue that the formal definition of PPGIS is 'nebulous' and generally inconsistent (Tulloch, 2008). According to Gregory, PPGIS may be sanctioned by government as a means to expand public participation and consultation (G. Brown et al, 2013). He also asserts that both VGI and PPGIS involve investigation and identification of locations that are important to individuals with the major difference being their respective purposes. Brown is of the opinion that PPGIS's main purpose is to inform planning and policy while that of VGI is for the participants' enjoyment. Others argue that the difference between VGI and PPGIS is more of an issue of semantics than substance (Hall et al. 2010). According to Hall both terms refer to the same thing while Brown views PPGIS referring to planning policy while VGI referring to pleasurable participation.

For purposes of this thesis, PPGIS is considered a component of VGI. In my opinion VGI consist of generic methods for creating and disseminating cartographic data. I assert that VGI is the broader term as it accommodates multiple motivations, including informing policy and planning, and not just for entertainment. Some of these contributions may be made by users who are compensated, as is the example of MapKibera. In my terminology, PPGIS could be considered the field that falls under VGI that allows for this creation and dissemination by different actors for the purposes of empowerment and enhancing public participation. It also includes this notion of actively engaging citizens with data to help break down knowledge barriers and is therefore considered more of a process than a tool.

The development of Google Maps in 2002 was a watershed moment for VGI. This is due to the fact that citizens could consume maps digitally over the internet bringing about the rise of what some scholars have termed the "geoweb". Moreover, the rapid advancement in different technologies such as third and fourth generation mobile networks, Global Positioning Systems and cloud computing have and continue to transform how geographic data is created, shared and disseminated. Thanks to Google Maps, users could not only search for traditional spatial or map based information but also other forms of digital information such as Wikipedia entries, photos and videos. The rise of social media has also made it easy for people to share location information, whether using manual technologies such as tagging to employing more advanced tools such as handsets with embedded GPS chips. As opposed to earlier technologies that made members of the public largely consumers of cartographic data, VGI brought about a shift in power from the traditional producers of cartographic data such as mapping agencies or statistical agencies to the ordinary citizen by empowering them to be co-creators of spatial data.

This change in power has resulted in several innovations targeting different uses. One simple example was in the rise of the term 'mash-up'. According to Darlene Fitcher, a mashup is a web application that uses content from several sources to create a single new service in a single user interface (Fitcher, 2009). She traces the term mashup to pop music, where people seamlessly combine music from one song with the vocal track from another thereby mashing them together to create something new. With the growth of the internet companies in early 2000, there was an increase in using data from such companies to create new services. The rise of mashups was largely due to the use of web mapping technologies in tandem with other tools whose main functionality was not mapping in one way or the other. For example, consider the popular Ushahidi platform which was used to document election violence during the aftermath of Kenya's 2007 election (Ushahidi, 2007). The Ushahidi platform allowed people to crowdsource a data set of citizen reports of violence and placed them on a map. This dataset was curated from online sources, text reports as well as social media feeds. These feeds were then geolocated and placed on a map, mashing up social media data with a geographic presentation. Another example was use of mashup functionalities in documenting the cholera outbreak experienced in London's Soho area superimposed with water supply lines then (Goodchild, 2007). These examples show how mapping is being redefined and continues to be used as a tool for social change.

Not only has the internet liberated cartography, the rise in the use of mobile phones has also resulted in broad uses of location based applications. The mobile phone has over the past ten years greatly increased people's access to the internet. According to a World Bank report, no other technology has been in the hands of so many people in so many countries in such a short period of time (World Bank, 2008). Increasingly, the rise in use and constant improvements to smartphones has resulted in a situation where one who has access to one of these devices can be considered a "citizen sensor" (Goodchild, 2007). Phones now are equipped with an array of different sensors such as accelerometers, gyro sensors and location based sensors such as GPS sensors. Additionally, phone based cameras used in tandem with embedded GPS chips help not only in collecting cartographic information but also make it possible for citizens to create accompanying hypertext and linked media, thus vastly improving the consumption of maps and map based products. Without the mobile phone, this perhaps would be increasingly difficult to achieve. As Goodchild asserts, there have been few tools that enable citizens to collect, verify and distribute such information (Goodchild, 2007).

The mobile phone perhaps has had a greater influence in emerging economies more than in well developed countries. To most people in emerging economies, the mobile phone is actually the primary device. It is the main device through which most access the internet and meet their other computational needs. The mobile phone is indeed a powerful tool for empowerment and development in such contexts. A great example is the use of mobile money accounts in Kenya that has vastly brought out financial inclusion to a large constituent population that was previously unbanked. Some of those who were previously unbanked, now through a mobile phone have access to financial services which they did not have access to before. It is estimated that more than half the adult population of Kenya now uses mobile money accounts vastly improving access to financial services (CGAP, 2015).

According to the Kenya Communications Authority quarterly statistics, Kenya attained a mobile phone penetration rate of 80.5 percent (CCK, 2015) in 2015 as opposed to the year 2012 where the mobile penetration stood at 69 percent (GSMA, 2012). There has also been a steady increase in internet penetration. According to the industry regulator, Communications Commission of Kenya (CCK), internet penetration as at May 2013 stood at 41 percent, in June 2012 at 35 percent while in September 2011 the penetration rate stood at 22 percent. According to their recently released statistics, the internet penetration stands at 57 percent as of February 2015. As depicted in the image below, mobile data subscriptions account for more than 50 percent of the total internet users. Few who access internet via computers through terrestrial networks and fibre optic cables account for less than 50 percent. This shows that more and more people are gaining access to the internet and most are accessing the internet through mobile phones.

Table 16: Internet Subscriptions and Internet Users

Internet/Data Subscriptions	*Jun 14*	*Mar 13*	*Quarterly Variation (%)*	*FY 2013/14*	*FY 2012/13*	*Annual Variation (%)*
Total Internet Subscriptions	14,030,036	13,356,415	5.0	14,030,036	12,432,308	12.9
Mobile Data/Internet Subscriptions	13,930,694	13,257,309	5.1	13,930,694	12,340,005	12.9
Terrestrial Wireless Data/Internet Subscriptions	17,169	16,540	3.8	17,169	21,282	-19.3
Satellite Data/Internet Subscriptions	646	700	-7.7	646	1,278	-49.5
Fixed DSL Data/Internet Subscriptions	12,129	12,547	-3.3	12,129	11,512	5.4
Fixed Fibre Optic Data/Internet Subscriptions	69,373	69,377	0.0	69,373	58,197	19.2
Fixed Cable Modem Subscriptions	25	25	0.0	25	25	0.0
Total Internet Users[4]	22,319,684	21,679,309	3.0	22,319,684	19,654,925	13.6

Source: CA, Operators' Re

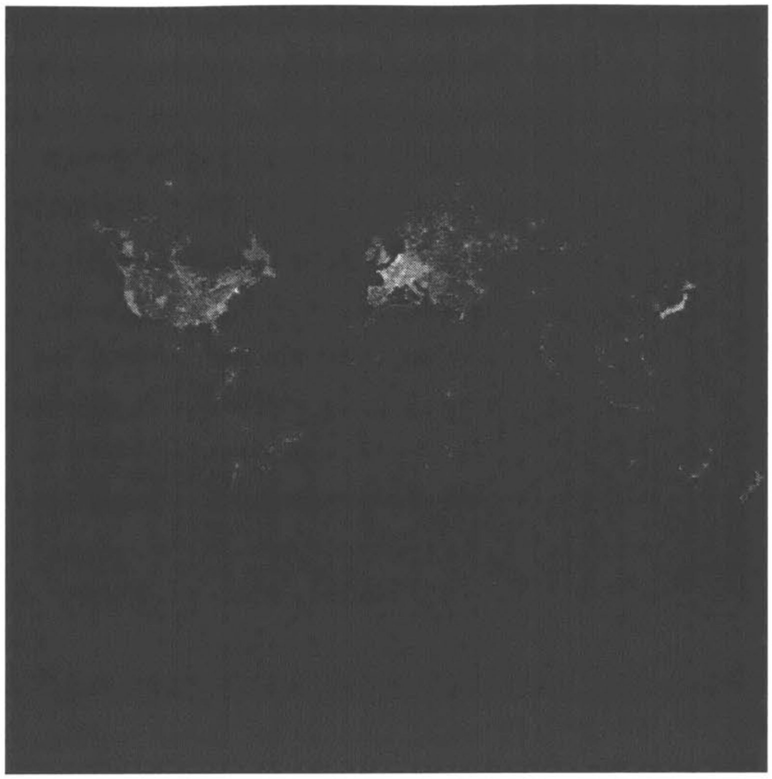
Despite such great numbers many citizens are disenfranchised simply because they do not have access to mobile phones. Of those surveyed in the RIA study, those who did not have access to mobile phones mentioned affordability as the main reason as to why they did not have phones. If we were to look at how this affects VGI, this affects not only the quality of contributions but also the nature of the contributions. This exclusion of local contributors with local knowledge in poor and rural areas results in poorer contributions mainly due to the lack of access to necessary devices and the technical skills needed to contribute. Moreover, there is need to look at this as more than a question of access to devices. As Hargittai rightly notes, there is need to look at the digital divide more than just from the element of access (Hargittai, 2008). According to a study done in Kenya on how those with phones use such phones, all respondents surveyed use the mobile phone to mainly make and receive phone calls. In second place, 82 percent of respondents interviewed use short messaging services while in third place 77 percent of respondents use cellphones to send and receive money. Only 25 percent of the respondents surveyed use the internet (RIA, 2012).

As Ezther Hargitai states in 'The digital reproduction of inequality', the digital divide is a term used to refer to the unequal distribution of the medium. One important element she highlights is that of access. Not all members of a specific community have access to the medium. A good case study of this is the internet. She further argues that looking at access to the medium is a very simplistic way to look into digital inequalities. A more refined approach considers different aspects of the divide, on details such as quality of equipment, autonomy of use, social support networks, experience,user skills in addition of different usage types(Dimaggio et al 2004,Norris 2001, Hargitai 2008, Dewan and Riggins 2006). These aspects of the digital divide affect participation in VGI anywhere in the world, but particularly in a developing nation like Kenya.

The use of VGI tools that require persistent internet connections is likely to broaden these digital divides. This is simply because in most of these emerging markets, mobile internet access is based on prepaid models. Subscribers purchase airtime credit and load the credit on their phones. Subscribers are then billed according to what they use. This is unlike in some American and European markets where subscribers purchase a monthly subscription with little or no caps on internet usage, where subscribers will not be affected costwise by persistent internet connections. Not only do such persistent internet connections make it more costly for individuals to contribute in emerging markets, but this cost divide also affects people's perceptions on spatial representations and identity.

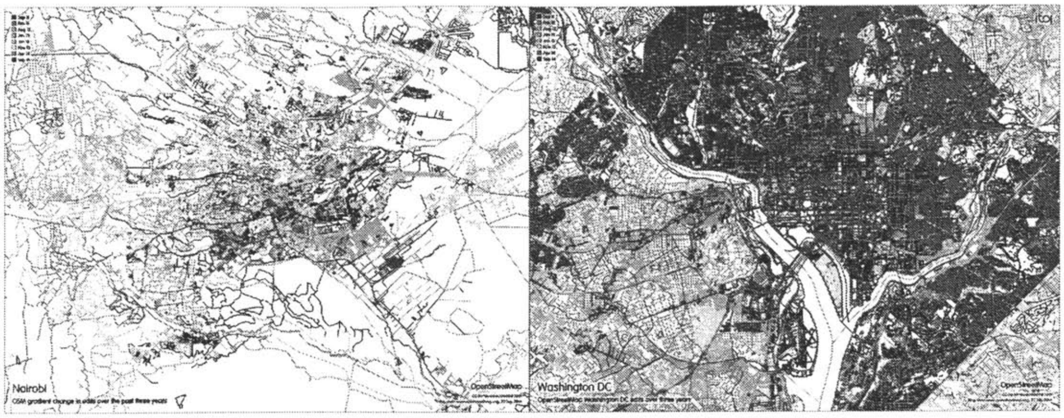
How can a question of access and affordability affect spatial identities and representation especially within the realm of VGI? When participants contribute their data to large VGI projects, they are motivated, in part, by seeing what they produce represented on a map. Different motivations are manifest in other online communities such as Wikipedia and other open source communities. According to Heng-Li Yang and Cheng-Yu Lai, those who contribute to Wikipedia are driven to contribute since they voluntarily contribute and the individuals' enjoyment is the greatest motivator. They also assert that other than this sense of enjoyment, contributors also feel a sense of personal achievement when they contribute (Cheng et al 2014). Contributors to OSM also have similar motivations, though at times enjoyment may not be the primary cause to participate, especially when volunteer geographic information is being used for PPGIS purposes. Ideological motivations such as unique ethos or having information being free to everyone is common to most online communities but what sets OSM apart from the other communities is this ability to define and share spatial representation and spatial identities. Questions of spatial representation shape motivations as to why contributors share specific data and see other data as inappropriate to collect and share. It goes without saying that barriers to viewing how individuals and communities are represented also becomes a barrier to how and what they can contribute. It is akin to coloring in the dark. Because of such architectures present in OSM, new forms of the digital divide have come about. More interestingly we need to examine how these new forms of the digital divide have affected and continue to affect contributions to OpenStreetMap.

It is important to look at how people have contributed to OSM over the past 10 years to understand questions of representation. The image below I created from OSM edits as at 2015 depicts contributions to OSM, showing the density of edits. What is more significant to take from the visualisation is that certain areas are more represented than others. The US, Western Europe and Japan are well represented in terms of edits, while sub-Saharan Africa, South America and parts of China and Russia are poorly represented. Are there reasons that could lead to this imbalance in contribution densities? What this image also brings to the fore is that some of the areas with poor internet penetration rates have fewer contributions.



OSM Node Density Edits 2015

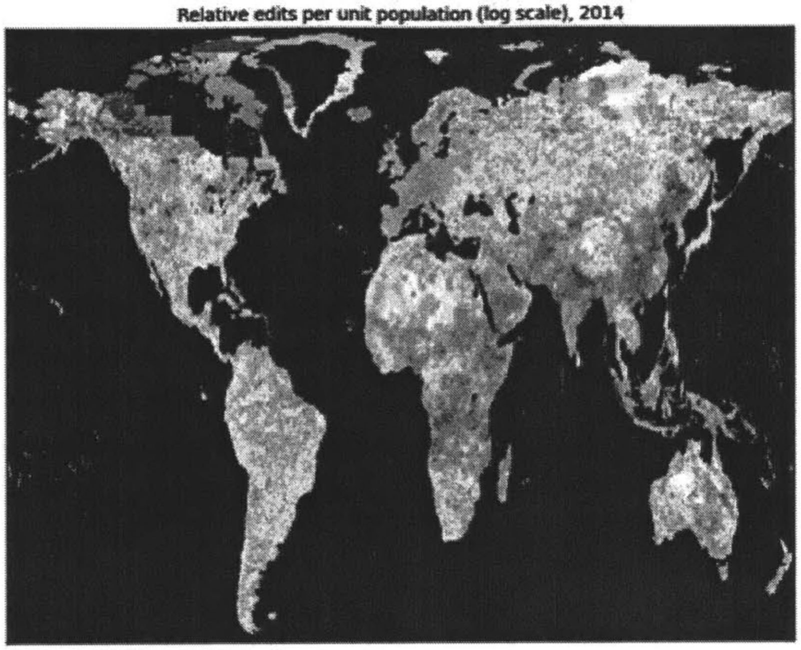
Another way to analyse contributions to OSM is to look for any relationships between population and contributions to OSM. This is as depicted in the next image. On comparison of edits made per population density, areas covered in blue are areas with high population densities and high edit frequencies. Areas covered in red are more concerning as these are areas with high population densities but very few edits on OSM. The image below shows the distribution of edits across two cities that lie on opposite side of this spectrum. Nairobi which lies in Africa and with a population of 3 million while DC on the other with a population of around 600 thousand.



Comparison of OSM edits between Nairobi and Washington DC

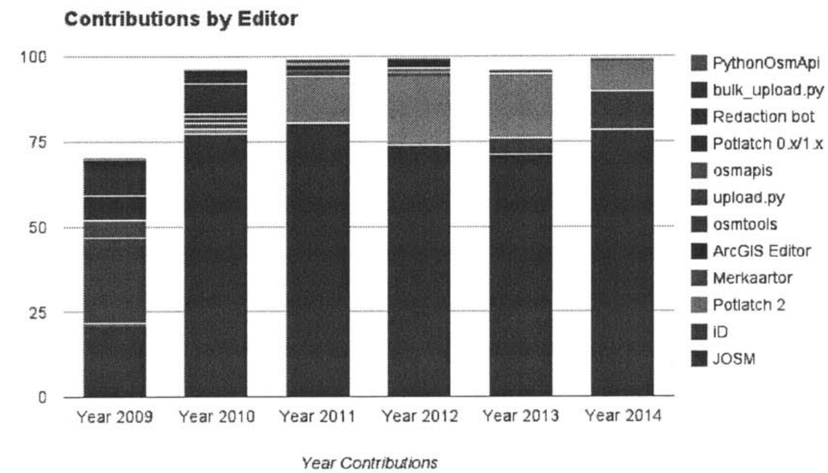
This indeed is a major cause of worry for those looking to empower such communities using VGI. More needs to be done to understand the discrepancy between OSM edits and population densities. Of course population is just one of the variables which could be studied. Other areas that would be good to look into include but are not limited to, internet penetration rates, the gross domestic products of the relevant countries, mobile phone penetration into these regions, amongst others. For purposes of this thesis, one way to answer this question is by analyzing the tools used to make edits to this global corpus of cartographic data.

Relative edits per unit population (log scale), 2014



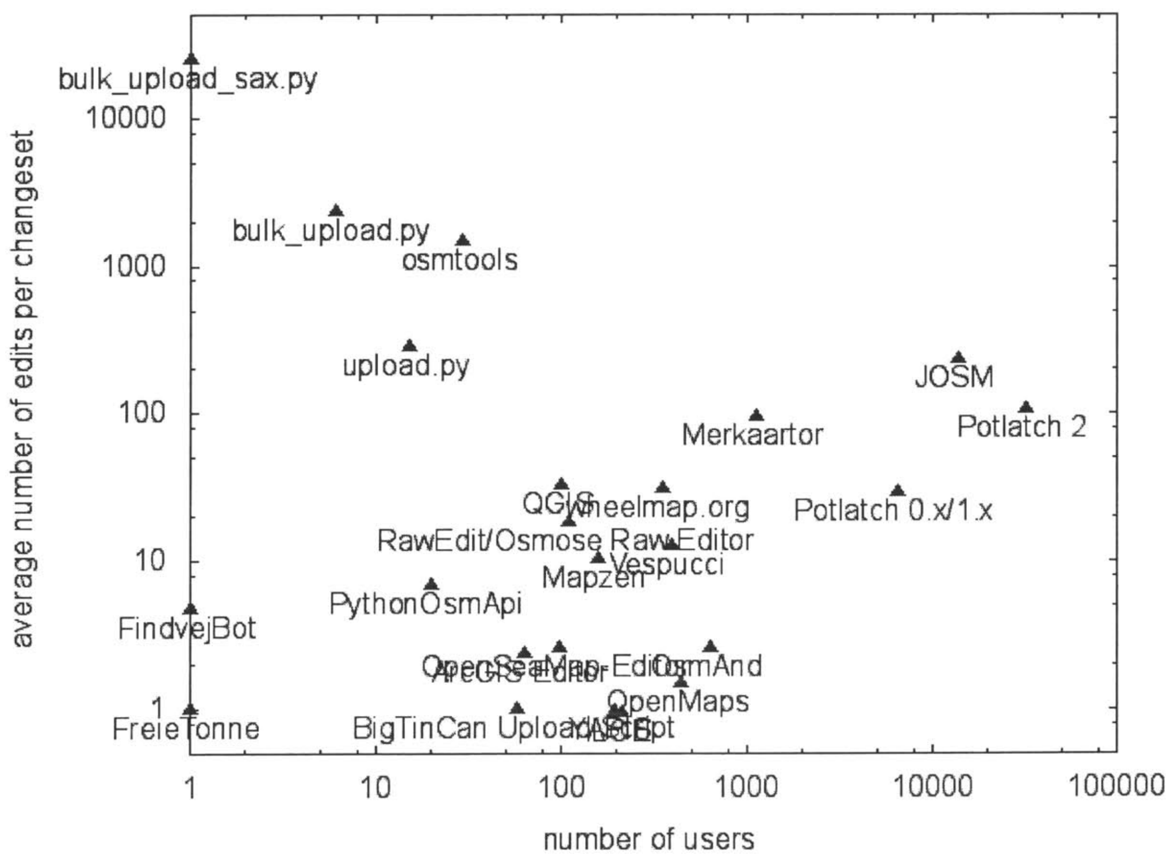
In analyzing edits made to OpenStreetMap by editor used, a couple of issues come to the fore. The chart below depicts tools used and the weight these tools have had over the years. Most of the tools used to edit OSM over the past years have been desktop and web based tools. Very few edits have been made via mobile phone editors. The top mobile editor to OSM has produced an average of 0.7 percent of all the edits made. Could there be a reason as to why very few edits are being registered by mobile handsets despite phones outnumbering all other devices? I believe there are two reasons for this disparity. One is that when OpenStreetMap came to being ten years ago, it was by design to be used over the web and not to be consumed or edited through phones. A look at how Wikipedia is architected also points to this possible cause. However, Wikipedia has made arrangements with some mobile service providers to make access to Wikipedia free. Sadly though such arrangements have certain political, social and economic implications. For instance, providing free access will affect people's perception of the internet. Initiatives such as Internet.org, are likely to not only mould perceptions on what is relevant on the internet but also provide a walled garden limiting peoples options to information. Provision of such services for free may not be a good approach for OSM or any other web service. It is also important to note that Wikipedia lacks mobile based editors.

Another important reason could be that mobile phones ten years ago were not as powerful as those that we have access to currently. To better answer this question, it is important to look at mobile editors that have existed to date.



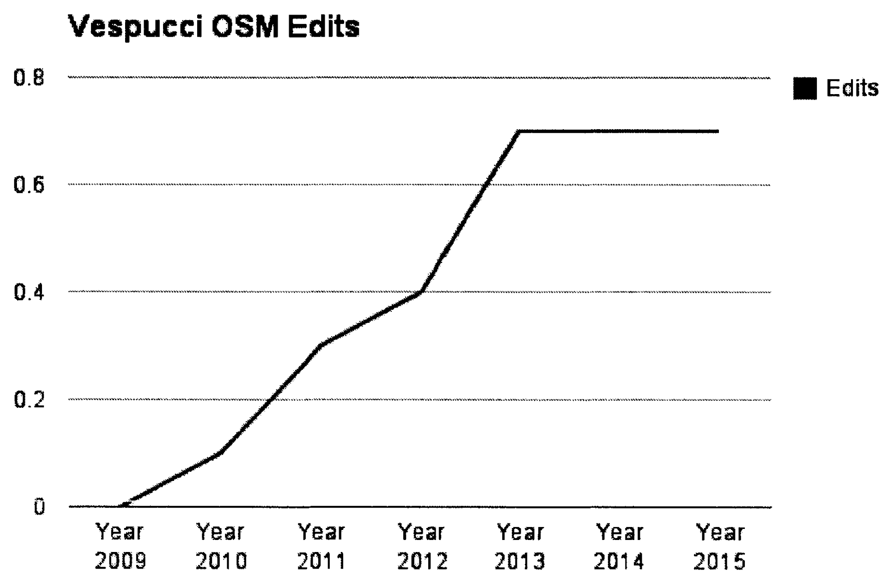
3 Perspectives on Existing Tools

As highlighted earlier, most of the edits made to OpenStreetMap are largely made using web and desktop clients. What mobile tools are available if any? It is necessary to evaluate mobile tools available and gauge their effectiveness and challenges they present when contributing to OSM. Tools evaluated in this section are mobile applications used to make contributions directly to OSM. It is important to note that edits from mobile applications being evaluated are responsible for very few edits within the larger OSM contributor dataset. Hence comparing edits made to OSM is likely to drown out contributions of such applications. Another mechanism would be to look at both contributor numbers and user population numbers. The scatter plot below shows editor statistics based on number of users.



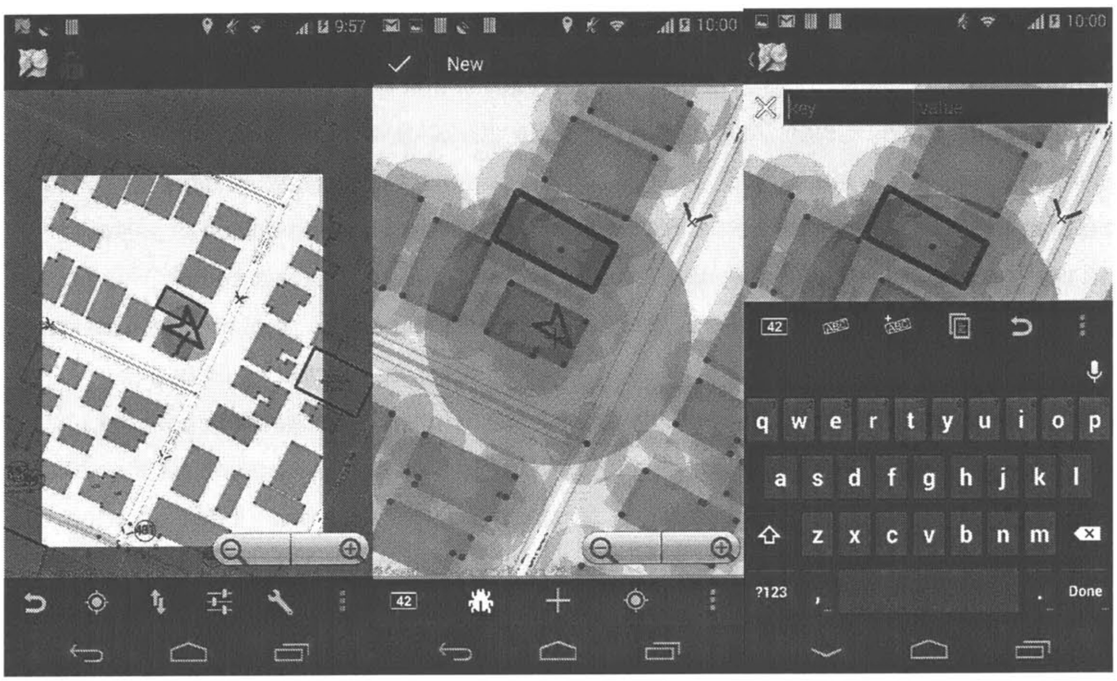
3.1 Vespucci

According to the OSM editor usage statistics, the Vespucci mobile application is responsible for 0.7% of edits made in the year 2015. The chart below illustrates percentage edits by Vespucci registered from 2009 until the year 2015.



Vespucci Edits from Year 2009- 2015

Since the launch of Vespucci, there had been a rise in edits being made to OSM through the tool, sadly over the past two years the application has not seen registered any growth in edits. This perhaps could point to two things, despite an increase in edits, most of those editing are using the current web and desktop tools. It is important to note here that the OSM ID web editor which was introduced in 2012 now commands the greatest share of edits to OSM. Perhaps with the increase in contributions, most are opting for the web based rather than mobile application editors. The other reason that could explain this is that perhaps more contributors who joined OSM were first exposed to OSM as a web or desktop environment first rather than through a mobile environment first. Contributors who are working from satellite imagery are more likely to be working from the desktop because of the size of the data and the size of their screens. People who make contributions from the field seem to be more inclined to use mobile. Based on editor use, both JOSM and the OSM ID editor account for more than eighty percent of edits made in the year 2015. This means most edits are being made by tracing and ground truthing using satellite imagery.



Screenshots of the Vespucci Application

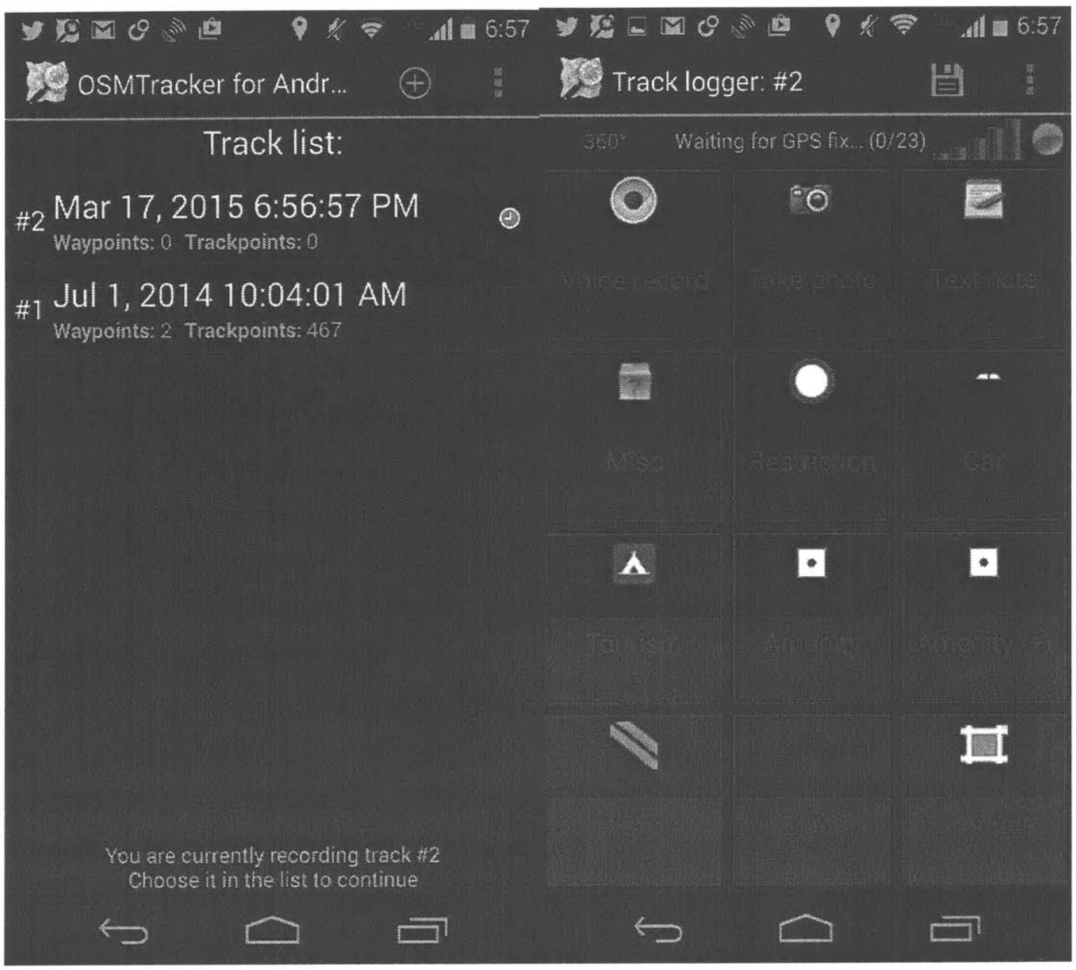
Vespucci is a great mobile tool that enables people to contribute to OSM. One of it's main advantages as opposed to the other applications is that is has a cache which enables one to perform editing on connections that are not persistent. The cache does get outdated and still has some dependency on internet connectivity. It is a cache that is time and zoom based. The cache is limited to the current view during use and at the zoom specified. This means that without the connection, elements become distorted and blurred when one tries to adjust zoom settings.

Another great feature is its clear demarcation between editing and consumption. One can easily view the maps using the tool, but if one needs to make edits, they have to press the unlock button. This is depicted in the first image of the application's screenshots as a lock button on the toolbar. Vespucci also has predefined categories that help in editing metadata hence making it easy for people to contribute metadata. This implies one does not need any knowledge of the underlying metadata when adding elements to the map. A simple example of metadata referred to could be store opening and closing hours, type of facility being mapped amongst others. While this feature may be attractive to users, it could also lead to chaos as lack of metadata standards frequently arise providing more work for moderators.

Another possible impediment when using Vespucci is that on editing, map elements are only limited to paths or linestrings. To illustrate this point, the Vespucci design assumes that any unit that is to be mapped is made up of a series of points connected by lines. In other words the base unit in mapping is a series of points interconnected by lined. Simple addition of points is not possible. The second image in the figure highlighted above illustrates this further. This poses a design challenge especially when users are interested in adding entities that may be best be represented as points on a map rather than as a series of points connected by lines. For instance mapping a tree on Vespucci is difficult as one has to draw several points perhaps taking the form of a circle, then labeling it as a tree. A simple click on the map and addition of a point could make for a better editing experience. Vespucci additionally makes it difficult when adding metadata since one is required to know the key and value pairs. This is because when adding metadata to already existing nodes one needs to does know the underlying OSM metadata structure. For instance an example of a key and value pair could be represented as the number of washrooms as the key and an integer for the actual number of washrooms present. In order to fill in such metadata one needs to know the key used within OSM in order to fill out such values. There is need for better metadata discovery when editing such data.

3.2 OSMTracker

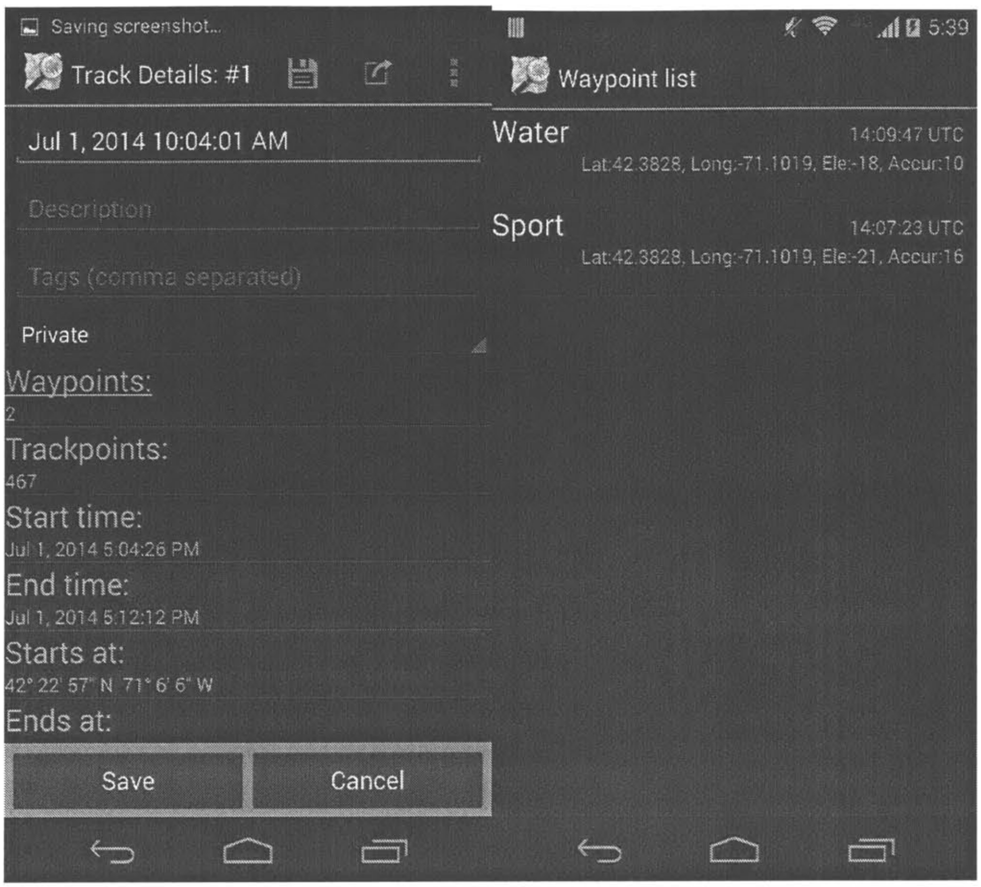
OSMTracker is an application that employs a different methodology in contributing to OSM. Sadly, there are no statistics available on how many edits have been made to OSM using the application since its launch. The application requires a user to collect points and accompanying metadata as they walk through a path collecting waypoints and measurements. Each waypoint provides the user with the ability to collect photos, GPS measurements and voice recordings. The track is then saved locally on the mobile phone and uploaded to OSM when one has a data connection available. The screenshots below paint a picture of how the application operates.



Screenshots of OSMTracker

The screenshot above show the listing of tracks and their status. As is evident on one of the screenshots, one of the tracks is in active mode, i.e., the track is currently being recorded. On selecting the track, one is able to enter different types of metadata. The screenshots below show track details and accompanying metadata. On selecting the waypoints that make up a track, one is able to see properties of the waypoints entered.

A track is made up of waypoints and each individual way point has accompanying metadata.



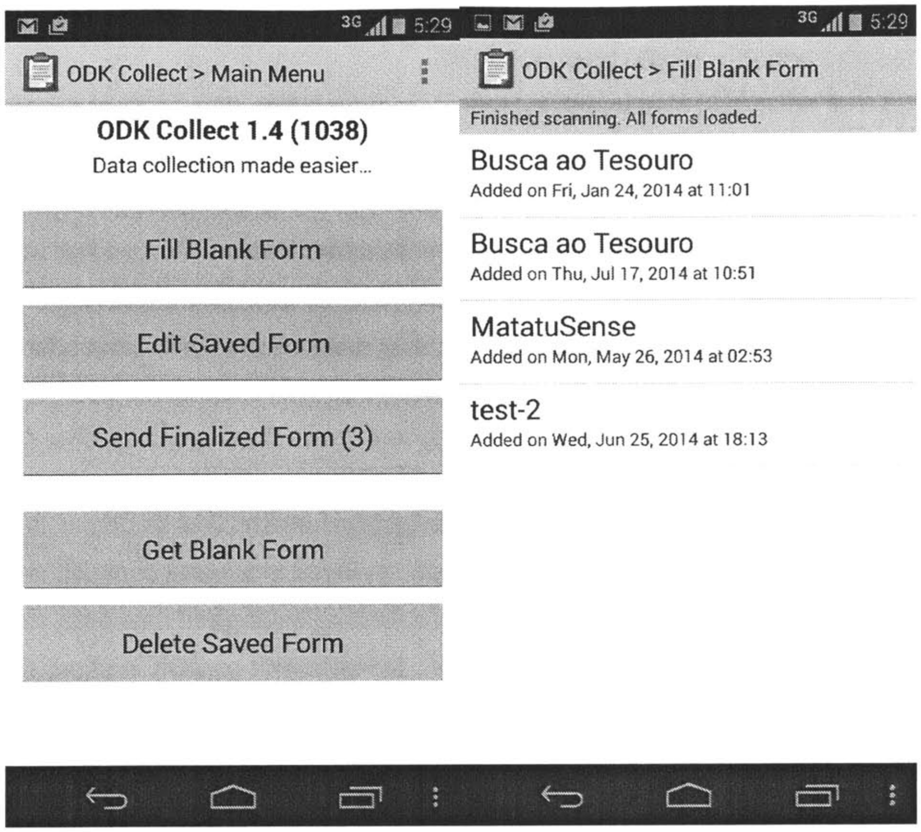
Screenshots of OSM Tracker

OSMTracker is a great application for adding data to OpenStreetMap. Its pervasive nature endears it to contributors since it is provides a link between one's activities and collecting data. For example a hiker can collect data as she participates on a hike, and could share that data with a community of hikers. One of the drawbacks by this application is that it fails to provide sufficient spatial context. This may lead to a situation where a user carrying out some data collection may not know what is already on the map and what is not. This makes it increasingly difficult to pinpoint forms of misrepresentation and generally identify items that are not represented and perhaps should be on the map.

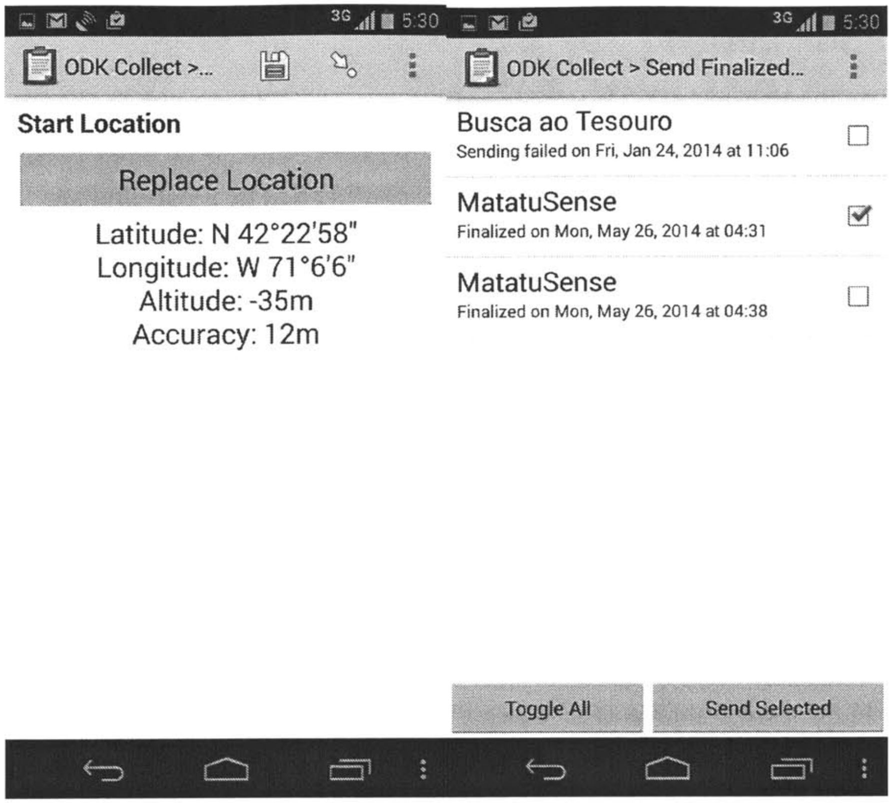
The fact that one could also use this application in an offline environment because one does not require a persistent internet connection when collecting data, is indeed powerful. A drawback of using this application is the lack of the user's ability to edit custom metadata. Its use of icons when editing metadata is powerful but limits user choices when adding metadata, especially metadata that is not represented by any of the icons available. That said, the editing of metadata through icons in OSMTracker is significantly more user-friendly than how metadata is edited in Vespucci.

3.3 OpenDataKit

OpenDataKit is a very popular mobile data collection tool that has its origins in the work of the University of Washington in Seattle. It has been deployed in several contexts and used in collecting and aggregating data. It has been used in collecting a range of data sets, from health issues to agricultural studies. As a mobile application, OpenDataKit or ODK in short, does not provide a mechanism through which one could send data directly to OSM hence it is not viewed as a primary OSM editor. Rather it is viewed as a tool where one could collect data, export the data to a machine-readable format and then using existing OSM APIs, upload the data to OSM. This is mostly achieved via computer or through some OSM importing web services such as the Humanitarian OpenStreetMap's(HOT) import manager. OpenDataKit is comprised of two main applications, the mobile application - ODK Collect - and a web service - ODK aggregate - whose main function is to aggregate data from multiple mobile phones to collect a data set. "We mainly use ODK because it is the only application that provides us with the ability to collect data in offline environments and sync the data later once verification was done", said Simon Kokoyo, one of the lead contributors to OSM and a project lead with the Map Mathare team, while I was conducting the evaluation for Mapjack with local communities. MapMathares is a team of slum residents who are involved in collecting data for different non governmental agencies in Mathare slum located in Nairobi.



OpenDataKit Screenshots



OpenDataKit Screenshots

One of the strengths ODK Collect affords contributors to OSM is this ability to easily design a data schema which fits within the application's user interface, providing a smooth and easy user experience in collecting data. This greatly reduces the barrier to contribute for those with little technological or mobile phone use experience. As with OSM Tracker, one main challenge about using ODK is its lack in providing spatial context. This results in situations where data collectors have no knowledge of what exists and what does not exist on the public map. The use of the optional mapping widget that ODK Collect comes with as a remedy to this does not solve this problem in many cases, since the widget requires a persistent internet connection. Using such a widget in certain contexts would lead to an increase in the cost to participate due to inherent internet connectivity costs.

4 Application Design

Based on the analysis of tools used to collect data for OSM via mobile phones, and past experiences working with communities to collect cartographic data, three main design interventions are proposed as mechanisms to foster participation by those who are disenfranchised. These three main design interventions include

- Offline mapping
- Custom layering
- Notifications for pervasive data collection

4.1 Offline Mapping

I realised the need for mobile platforms that make use of offline mapping and custom layering techniques when working with members of a community in Embu and environmental activists in Liberia in 2013. While working with Silas Siakor, who was awarded the Goldman Prize in 2006 for his work in conserving forests in Liberia, he mentioned the need of using mobile tools which were secure to collect data specifically around natural resource appropriation. Largely the initial motivation was to provide a secure mobile application which would help Siakor and his colleagues document illegal activities or whatever the team of activists attached to his organisation, Sustainable Development Institute, would find important to document. Data collection involved using Android smartphones with GPS devices. The application makes use of an encrypted local database and uses Tor libraries when syncing with central servers. Apart from location data, activists would record audio, video as well as images. The TIMBY (This is My BackYard) application has grown since then and had now been deployed by Code4Africa and currently being used by several citizen journalists across the continent.

Despite providing strong security features, allowing users to save data on their encrypted SD cards and sync when need be to a secure server through a secure connection, having a sense of spatial context when recording location seemed to be very important to the activists. We deployed a later version that made use of Google Maps API and would require mobile airtime or mobile internet to be able to view what they were recording on a map. Given the high cost of internet in Liberia, the activists protested against this new functionality.

Eventually, the embedded map was deprecated. Here was a serious problem: how do we provide spatial context in data or local applications while at the same time reducing the cost to access such mapping services? More needed to be done to provide offline maps. This is why during the design process, I arrived at the need to have offline maps to provide users not only with spatial context when collecting data but also as a mechanism to cheaply consume map based data.

4.2 Custom Layering

The Liberian activists also voiced the need to have custom layers which they could reference when adding data to the map. To provide an example, SDI as an institution had cartographic data representing the zones allocated to different resource extraction companies. While testing data collection, activists felt the need to be able to compare what they identified and to cross check with the map to see if extractive activities were legal. This came about when one of activists was walking around one of the dense forests in Liberia and noticed a truck belonging to a specific company taking part in logging operations. Given the vast and expansive nature of these forests, the activist had no spatial context or orientation to know if the company firstly has been granted access to the specific resource extraction blocks and secondly if the area is one of the resource extraction blocks where logging is allowed. To be able to make a decision regarding the specifics of such a case, activists need offline maps with custom layers.

Why use custom layers? Why not add these resource mining blocks to OSM as map features? This is mainly because of two reasons. The OSM architecture would require the activists to ask or petition the data working group to create new schema types for such mineral extraction blocks. As with my experience trying to get Kenyan shrines on a map, there are likely situations where the data working group would not allow for such a dataset, deeming it to be a local, not global, need. The other aspect is that even though these layers would be added, they are not likely to be represented on the global map. It is simply not possible to visualize all types of data because it would lead to a global map that could be cluttered and perhaps fall into the situation where these layers would affect how people in other areas such as Europe and the US want to be spatially represented.

4.3 Notifications

The design principles around the use of notifications was based on the work done in Center for Civic Media at MIT on Action Path. Notifications is a technical mechanism where small specific alerts are sent to a user's mobile application informing them or prod them for their attention. Action Path is a location based tool for civic reflection and engagement built by Erhardt Graeff (Graeff, 2014). How can we use persuasive technology to improve the number of contributions to OSM? During the design of the Mapjack, notifications were seen as an avenue where persuasive computing would be used to get more people involved by providing them with micro data collection tasks(Fogg, 2002).

The above three design interventions do not represent a perfect solution to mapping challenges in the developing world. Given the application is based on the Android platform, there is a huge constituent of the population in emerging markets where smart phones are not in use. What happens to those who have basic feature phones, phones whose main functionality is basically calling and texting? I explored this idea through work on PublicPapers which is a platform that seeks to empower communities with no access to smartphones. PublicPapers uses hand-edited paper maps to enable contribution to geographic data sets. PublicPapers is build on the work by Stamen on Fieldpapers (Stamen, 2015), the main difference being the use of custom grids which could be used for georeferencing using short messaging services. Users could add features to the map, then submit their changes using the georeference coordinates, transmitting information to OSM via mobile phone. This offline mapping also helped in designing the Android offline mapping experience for Mapjack. When people write on paper, there are no constraints as opposed to adding points through a mobile phone where one needs to work within the confines of the application's architecture.

Other design constraints involve the updating of the offline maps. Given one is operating in offline environments, data that has been collected can easily be sent to OSM when the phone has connectivity. As users are making edits and pushing them to this global OSM database the local copy of the data providing the offline maps needs to be updated as well. Given the nature of the updates and the general file size of an update, this can be a deterrent as it can make the map download costly. This is especially worrying for cities with large OSM concentration. Luckily some of the cities that are in emerging markets do not yet have large file sizes that would take long to download: the full city map for Nairobi currently is 3 MB, while the map of Washington

DC, a much less populous city, is 10MB. A possible solution I considered but did not deploy was a local mobile phone based versioning system, that operates very similar to git(the open source versioning system) where rather than updating the entire map, the map is unpacked and modified based on read and write header information to determine data that is to be merged into the raw file. The raw file is then updated and packed back into the minified form to fit on the phone.

The lack of open libraries to help in manipulating OSM's default compressed format was another obstacle. As mentioned earlier, developing these libraries was outside the scope of the thesis, as early investigations reveal that a lot more time would be needed to create open libraries to manipulate the PBF (Protocolbuffer Binary Format) format which the offline rendering relies upon. Another constraint related to the OSM compressed data format is that mobile computing still is not capable of certain computational processes. Lower level smart phones with poorer microchips are likely to too much time during unpacking, updating and packing the data back to the OSM data format. Measures taken to deal with some of these constraints include designing the application in such a way that there will be no need to unpack and repackage the offline maps. This is still a pending issue which could be addressed in future work.

5 Development Plan

Mapjack has been under development over a period of eight months. Initially the pieces envisioned were, a mechanism to render maps offline, and a mechanism to provide a git based data versioning system to preserve data edits and keep offline map data in sync with global OSM updates. Git here refers to the popular open source distributed version control system. Over the period, more features were added to improve user experience and to meet the needs to citizen cartographers. Ultimately, the three main pieces for this thesis were a mechanism for offline mapping, the use of layers to enable collaborative map making outside the bounds of the underlying base map, and a notifications system to provoke users to provide map or layer updates based on their location. The notification system is also backed by a web service where edits are pulled and push notifications published.

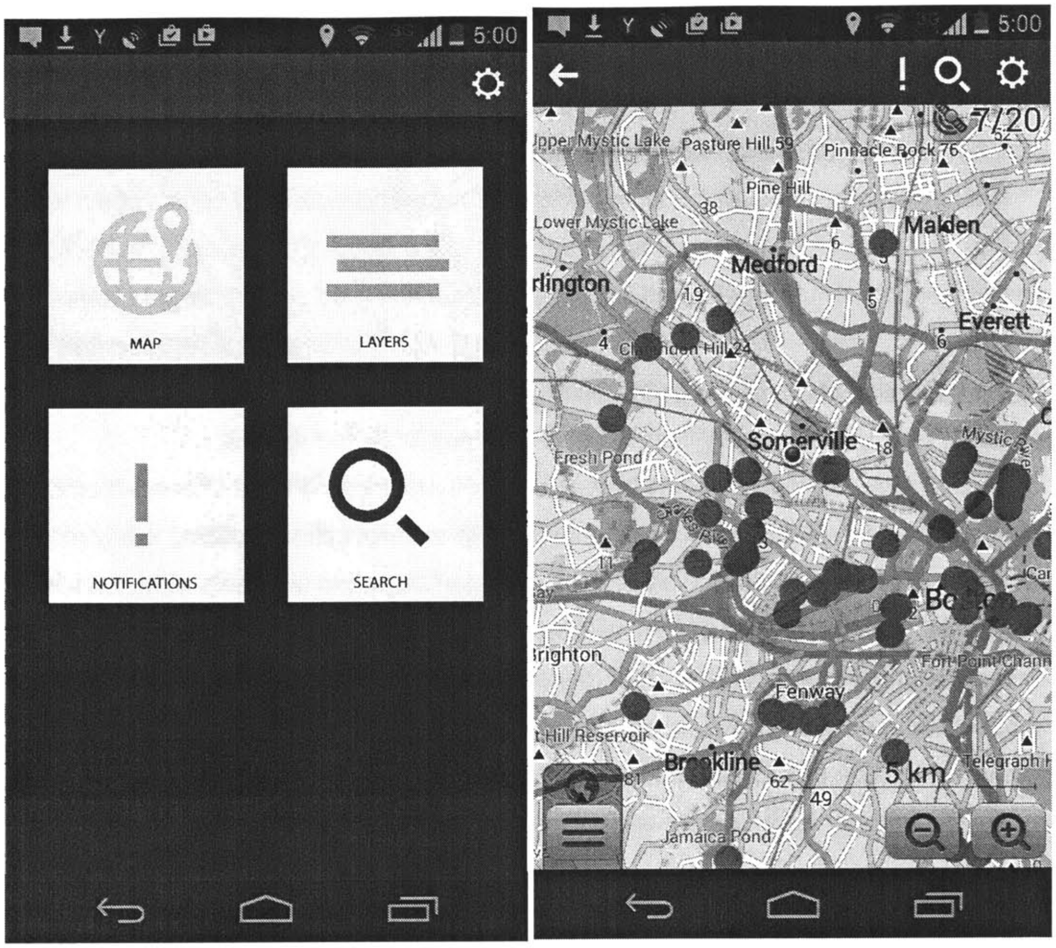
As part of the development process, the Android mobile operating system was chosen due to several factors. Most of those who have access to smartphones in emerging markets are on the Android platform. Secondly, Android as a mobile operating system is open, thus ideal when dealing with several open libraries for data and graphical manipulation(GSMA, 2014).

The offline mapping platform was the first piece to be built. This involved making use of OSM's data compression formats. It was challenging to decide which was the ideal format to use. The default OSM XML format is easier to manipulate due to the ease of dealing with XML. This would be important in the event users needed to view node metadata. For example, if this format would have been used, on clicking the map, one could view the tags and accompanying metadata for that node. Sadly, the native OSM XML format leads to large file sizes, which are difficult to use on mobile phones. By comparison, the compressed format is easy to manipulate and render, though the ability to perform queries on the fly is greatly hampered. The need to uncompress the entire compressed file and query a large xml document on the mobile phone on the fly could prove to be very time consuming (depending on hardware speeds) thus hampering the contribution experience.

In terms of user experience, the main motivation behind building offline map capabilities is to make map consumption easier and cheaper, therefore making it cheaper for citizen cartographers to participate. The cost to participate is reduced by making use of less data connectivity. Another underlying motivation is that since no tools exist that provide cartographic context, offline mapping and rendering helps provide context when participating therefore helping citizens in determining their spatial identities.

Custom layering was also built to address issues where citizens could collect whatever kind of data they wanted, whether or not it was part of the larger OSM project. In section 4.1, I expounded about activists' needs and interests in collecting data around illegal mining. Data such as the mining concession blocks is not data that would ordinarily be represented on the underlying OSM basemap. By providing such actors with mechanisms to collect data they are interested in, the more the underlying base map is also likely to improve.

Tackling this problem thoroughly may require work beyond what I have accomplished with Mapjack thus far. Rather than limiting layering to an application, building a tiny mobile based mapping service would directly address this problem, and also help other mobile mapping tools. With the advent of map servers, such as GeoServer, ESRI's ArcGIS Online, MapServer amongst others, a version of any of the above tailor made for the mobile phone would bring about a mobile based map publishing model that could improve all other map based applications. A mobile based map service could be used to publish a terrain map for instance or some satellite imagery for a specific area. The service would then allow one to use the satellite imagery being served by the map server on say the application Google Maps as a custom map.



Mapjack Screenshots

The notifications components was intended to send a broadcast message to people with the application installed to perform small contribution tasks. This was based on the study of Notes within OSM. OSM Notes are comments by contributors about errors or missing data on the map. Sadly more than one year since notes came into being, some notes have been largely left unattended and some unresolved. These notes often point to a contributor's displeasure around elements of misrepresentation. One mechanism to solve this would be to use geofencing and an alerts system to ask those around certain areas to collect the data in order to correct the map.

One of the critical components not completed at this point is the web service. One of the main functions of the web service is to act as the centralized repository of all edits that have been submitted by citizen cartographers. All edits made on individual citizen handsets need to be aggregated and merged to the underlying OSM basemap. This web service was also envisioned to act as a mechanism to give citizens offline map updates. During the initial mobile application setup, the web service would send the compressed OSM file requested and keep tabs of the map and database state when the download was requested. This would entail tracking the date, file size, area or areas downloaded, and a timestamp. This would then be created into a hash function. When syncing collected data, this hash is then sent to the web service which would then determine if the offline map needs an update. The user would then be prompted to update their offline map with the newly compressed OSM dump file. Another reason for the web service was as a mechanism to publish a layer to several devices and also act as the repository for the notifications. This work remains unfinished.

In summary, Mapjack is a mobile application built on the Android platform that enables citizen cartographers create, share and disseminate cartographic information. The application makes use of offline mapping techniques in an effort to reduce the cost to participate for citizen cartographers. It also employs the use of custom layers and notifications as an effort to bridge the divide in cartographic production.

6 User Experience

Several design interventions were used to tackle this question of under representation and how inclusive participatory mapping could be. The Mapjack mobile application is a simple application whose main objective is to make it cheaper to consume OSM based mapping data while also providing a tool that makes it easy to contribute. Some of the user design interventions include the use of status screens, offline map rendering, and custom layering. The next few chapters will highlight some of these interventions.

6.1 Status Screen

Status screens are very important in any mobile application. Letting users know which processes the phone is undertaking in the background not only helps keep users in the know but also provides them with information that sustains interest in using the application. Due to the nature of loading the cartographic offline dataset, which can take some time, there is need to keep users informed of libraries being used, their respective states when loading tiles and when rendering the very same tiles. Given that Mapjack is operating in an offline environment providing these updates is fundamental to the general experience in consuming data.

6.2 Single Page Form Design

The use of two main forms to collect point based data and collect elements of misrepresentation were other key interventions in the application. Addition of points of interest (POI's) was a mechanism to allow users to view the map in this offline environment and add edits to the map. The form used to enter data was limited to a subset of data. The metadata types include:

- The name or title
- The category
- Comments
- Addressing information

As opposed to most of the other applications, there is no provision for additional key value pairs for additional metadata information. This is by design in order to make it easy to collect data. Additional data can then be sourced through other means. For purposes of this thesis and most of the communities I was working in, key value pairs for metadata make data collection strenuous and were left out in entering details about the metadata.

6.3 Offline Mapping

The provision of offline maps was in itself a design intervention. The ability to render a base map and have users focus on specific base maps helps to increase not only consumption but also contributions. The offline map is rendered on the fly and may take a while to render depending on the density of edits. Another design intervention was rendering based on the GPS's current location and a general zoom of 13. This is simply because when one is making edits on a mobile phone with a GPS chip, the default should be their current location and there is need to only render what is relevant based on location. This reduces the processing and rendering time significantly. On zooming out, render times take longer, largely because of the extra processing required to handle larger spatially referenced areas and and because the geographic extent has increased thus rendering more feature sets.

6.4 Custom Layering

The provision of custom layers enables users to have the freedom to choose what they can share with a global audience or simply share within specific groups. It is important to highlight that the principles behind the provision of offline base maps are the same as dealing with custom layers. Custom layers simply provide flexibility which the wider OSM dataset can not provide. Having custom layering presented as part of a different user experience helps users have a mental differentiation between what they share publicly and what they share with closed networks to which they belong.

6.5 Search

During the deployment of the tool, one major request by users was the need to provide search. Provision of search from compressed data was indeed difficult. The ability to search for entities within the map or better yet search for certain points of interest not within the OSM ecosystem but within their private datasets is likely to help improve users efficacy when collecting data. For example, a team of educational officers attached to the Mombasa state government may need to add a school or need to update information about the school. Given the great underrepresentation of several amenities in Africa, the school may be missing on this global OSM public dataset. Subsequently, the state government may have within the phones provided to the officers provided the list of the schools within their area of interest of operation. To provide a better experience in cross referencing such data, the custom datasets within the phone could be searched and any results would then be presented to the user. The user would then update the dataset or alternatively decide to also update the OSM base map. Being able to provide extensive search where entities that exist within local layers but not the OSM dataset could encourage such users to contribute what they have locally to OSM. At the time of writing this thesis the search functionality was not implemented.

6.6 Android's Material Design

The use of Android's material design principles are also some of the design principles used to make a smooth user experience when consuming and collecting data. These principles were introduced in the year 2014 and hence most of the tools mentioned above do not use these principles. Some of these principles include using navigational bars that make it easy to navigate through mobile applications. The material design principles also help developers build applications that emphasize user context. Providing spatial context is important in collecting cartographic data, aligning spatial context with some of these principles of material design are likely to improve the general user experience.

7 Challenges

Several challenges both technical and administrative were encountered while developing and deploying Mapjack. Despite these challenges, different efforts were employed to mitigate some while others proved to be too difficult to be solved within the given time period. Some of these challenges are highlighted below.

7.1 Community Organizing

In order to evaluate the success of this thesis, there was a need to perform some tests with specific communities. Reaching out to some of these communities was indeed difficult especially given the eight hour time zone difference and the challenge in identifying mutual areas of interest between some of these communities and the affordances of the tool. I initially considered working with the group of activists in Liberia I had collaborated with earlier, however performing a deployment with them was made impossible due to the global health concerns and the Ebola crisis that affected the country at the time. Working with the other community group based in Embu, Kenya, whose focus was on biodiversity was also not possible as most of the community members who worked with the group four years ago moved from the location where this community was based, as with time, group dynamics change.

Identifying communities who have data collection as part of their processes also proved to be difficult. This is simply because groups that already understand the need for data or use data collection tools are ideal since they already have some preliminary understanding around data, its collection and its processing. Using communities with no prior experience in such skillsets would result in spending more time training members making it very difficult to deploy the tool in the given time frame.

Given that such deployments take time and have to take into consideration community members schedules, settling on dates and getting a consensus on ideal dates to carry out deployments takes a lot of time and effort. Weaving through the different interests between community organizers and members is a process that usually has to be handled with extreme care in order not to disenfranchise certain members of these communities. Given some of these members are also taking time off their busy schedules, financial remuneration is also an issue.

In this case, since this is a research project, it was made clear that no financial or any other forms of financial remuneration would be availed to participants.

7.2 File Handling and Compression Formats

Working with the default OSM data compression format, the Protocolbuffer Binary Format (PBF) presented some advantages as well as disadvantages. PBF is primarily seen as an alternative to the XML format that OSM uses. This format is ideal when dealing with large OSM datasets as it is five times faster to write and six times faster to read than a compressed OSM XML data file. No only does the PBF afford us faster processing times, it is also ideal when rendering maps in an offline environment. This is because mobile devices do not have as much storage space as conventional computers, hence PBF saves space while at the same time providing an easy mechanism to read and write to the PBF. Unfortunately, the PBF has some drawbacks. While building the tool, the PBF was great during rendering of the offline map. According to the intensity of map elements, this affected how fast rendering was done on the screen. Rendering a large area such as Massachusetts indeed takes slightly longer to render than say Nairobi. This challenge is not such a big issue since there is a direct correlation between map elements and rendering speed. A building will take a shorter time to render as opposed to all buildings on a university campus.

The lack of Java-based manipulation libraries for PBF makes it increasingly difficult when working with the PBF format. Take the example of when one clicks on specific map elements to obtain metadata stored in the PBF. This metadata is stored in the PBF as key and value pairs. Being able to query metadata is important in order to inform data collectors on what metadata is missing or needs to be changed, very similar to how the offline map is able to provide data collectors with spatial context when collecting data. Sadly at the time of developing the tool, no such library existed that could help in both reading and writing. Efforts to reuse some of the legacy Java based projects such as Osmosis, took longer than anticipated. Osmosis is a Java based command line application for processing OSM data and is the primary PBF reader and writer. Incorporating Osmosis would involve translating most of the Osmosis functionalities to an Android environment to provide custom helpers to access such functions when reading and writing to the PBF. As I mention later on, this is an area where the tool could be improved and could form part of future work.

7.3 Custom Metadata Schema

Given OSM's storage of metadata as key and value pairs, providing an intuitive interface where users have the ability to input their own custom key, value pairs was necessary, but proved to be challenging. This was because since metadata fields for some of the elements are stored in an sqlite mini database. Thus there was need to develop a mini form designer where entry of such key value pairs would occur. The other impediment was that when adding edits, the mentioned key value pairs would then be part of the form elements when adding metadata. Several applications such as OSMTracker make use of two text fields where one could enter key, value pairs. This unfortunately brings about data inconsistency especially since the key may not be constant across all edits being made. For instance, perhaps a data collector was to collect points about schools. Part of the metadata could be the number of classrooms. A simple case could be the key being the text "classrooms" with the value being an integer. During the course of adding metadata, this process not only would be tedious since one has to constantly enter the key "classrooms", but there is need to identify already existing keys and create better form entry experiences when adding additional points and their respective metadata tags. This means when heading to another school, one should not have to fill the same key over and over again.

This custom metadata schema also presents a challenge when working with large communities. Given they are operating using mobile devices, the likelihood of errors in tags is high, which creates problems when syncing these key value pairs to OSM. There is a need to have a central repository where all these edits are moderated and errors rectified before committing such data to OSM. Pushing edits to OSM without some form of quality control or gatekeeping will also affect consumption and querying of such data. For instance, if one was to query all schools where the 'classroom' tag value is greater than five and the metadata pushed has some inconsistencies such as a "classrooms" tag, results are likely to be inaccurate. Unfortunately for the first test, the schema was hard coded with common categories and tags to prevent such issues.

8 Evaluation

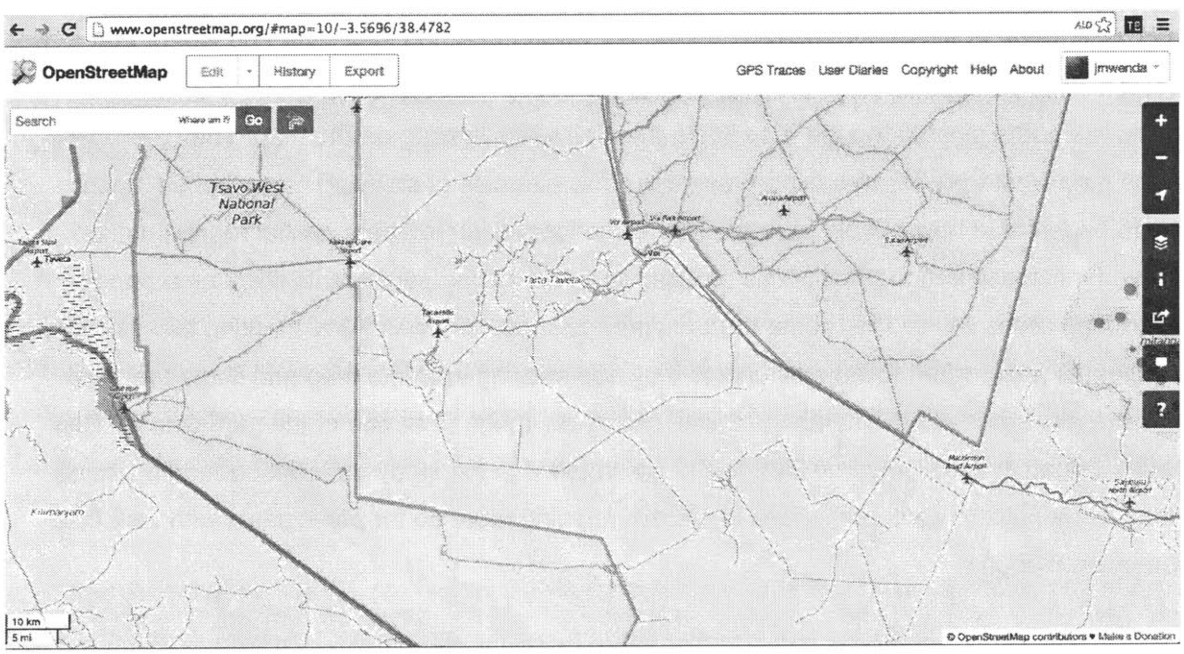
8.1 Overview

The principle objective of carrying out the evaluation was to determine if the use of Mapjack lowered the barrier to participate in creating and sharing cartographic data. It also sought to evaluate on the ease of use and consumption of cartographic data in offline environments, specifically via mobile phones. In order to evaluate the suitability of the tool, workshops were held with different communities who had the chance to use the tool to collect geographic information of interest. Interviews were then carried out with each of the participants after testing out the tool in real world situations.

8.2 User Selection

Mapjack was tested by three different communities during January 2015. The three different communities made available a total of thirty seven members. The communities where the tool was deployed were MapMathare, Swahilibox and Tsavo wildlife conservancy, all located in Kenya. MapMathare is a community made up of young civic actors who are interested in social justice in the slum area of Mathare. Mathare is the second largest slum in Nairobi after Kibera. Members of MapMathare frequently undertake data collection activities and they have extensive knowledge of OpenStreetMap. They have partnered with other non-governmental organisations to collect different kinds of data and add the data to OSM. Looking at the map of Nairobi, Mathare is well represented due to the activities most of these youth take with the support of non governmental and international donor organisations. Most of the members have more than three years experience in working with OpenStreetMap. Out of the thirty seven people surveyed, MapMathare produced three community members to test the tool. Unfortunately, due to time constraints other members of MapMathare could not make it to participate in the study.

The Tsavo Wildlife community is made up of members around the Tsavo National Park, a massive national park in the southeast of the country. This is a local community engaged in empowering local citizens to be part of environmental conservation. Tsavo National park is one of the few conservancies with a largest elephant population and was also home to the renowned man-eating lions. The community group invokes the help of local citizens in documenting activities such as environmental monitoring, documenting cases of human-wildlife conflict and decreasing carbon footprint by providing alternative energy sources. The main reason why this community was ideal to deploy the tool with was because the communities in this area are grossly underrepresented on OSM and are in areas where access to internet is very low or non existent. This area has no telecommunication network. For one to receive a call, one has to frequent certain hills or the nearby town center in order to obtain cell phone coverage. Most of those interviewed have never touched a computer but had mobile phones at the workshops. The image below shows the level of under representation. Members employed by the conservancy are also involved in data collection using paper based forms. Given their history in collecting data, this was also an ideal community to evaluate the tool with.



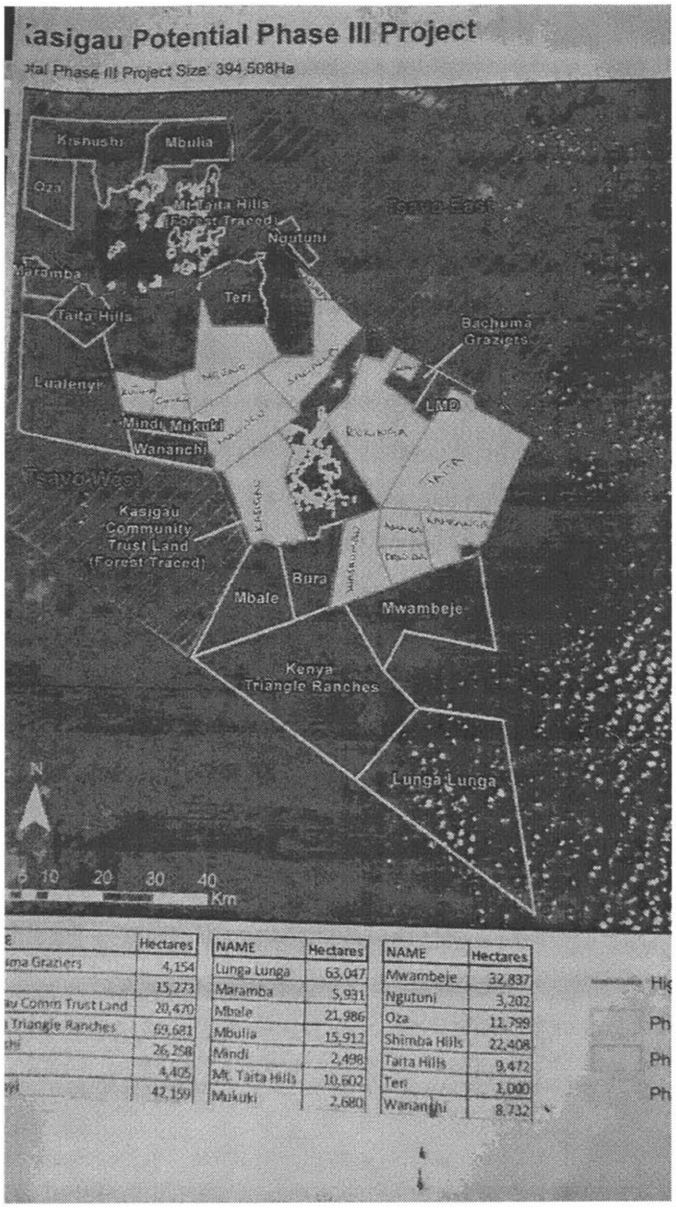
Tsavo Area Conservancy on OpenStreetMap

Swahilibox is a new tech member collaborative where people involved in Mombasa's tech scene gather. Mombasa is Kenya's second largest city. The collaborative now has 200 members and is growing. Most members involved in this study have been involved in mapping in one form or the other. All who participated in testing the tool had prior experience in using Google Maps and had used Google's Map Maker, which is Google's tool for users to add data to Google Maps. Most of those who were testing the tool had never used or come across OpenStreetMap. The main reason behind working with this group was their urban setting and their experience in using another competing mapping platform. It was also to provide a basis to learn from their experience in contributing to Google Maps and if the same principles apply to OSM.

8.3 Experimental Procedure

8.3.1 Workshops

The deployment process involved having workshops to first understand some of the processes some of these communities undertake when collecting their data and where they collect it. This process involved printing out papers based maps of the respective areas and having members pinpoint on the map items they felt were missing on the map. They then drew on the map what they felt was misrepresented. This exercise of studying paper based maps was to understand how people perceive maps and looked into matters spatial representation and under representation. Part of this process involved having participants mark on a paper based map areas where they come from and hence know very well. After marking these areas, participants were asked to list down items they see missing from the map and items that they think are misrepresented. The image below shows an image from one of the participants. This greatly helped in assessing how those who participated in the study understand reading maps. It also helps put into context the idea of citizen mapping more so for participants with very little literacy on mapping.



Paper based Map on Spatial Representation

After listing down items, members were asked to place those items in the form of a priority list. Thereafter, I facilitated a conversation with members of the various communities on what they deemed important to map. "Having the animal corridor represented on the map is important to us so that we can know how to prevent cases of human wildlife conflict", said one of the participants. "This map does not show any of the natural features present, yet they are very important", another participant mentioned. We then grouped items members listed into categories and listed them based on importance. This was to better understand what members of these communities perceive as important to be spatially represented. The image below shows the categorization of items based on importance for Tsavo Wildlife participants.

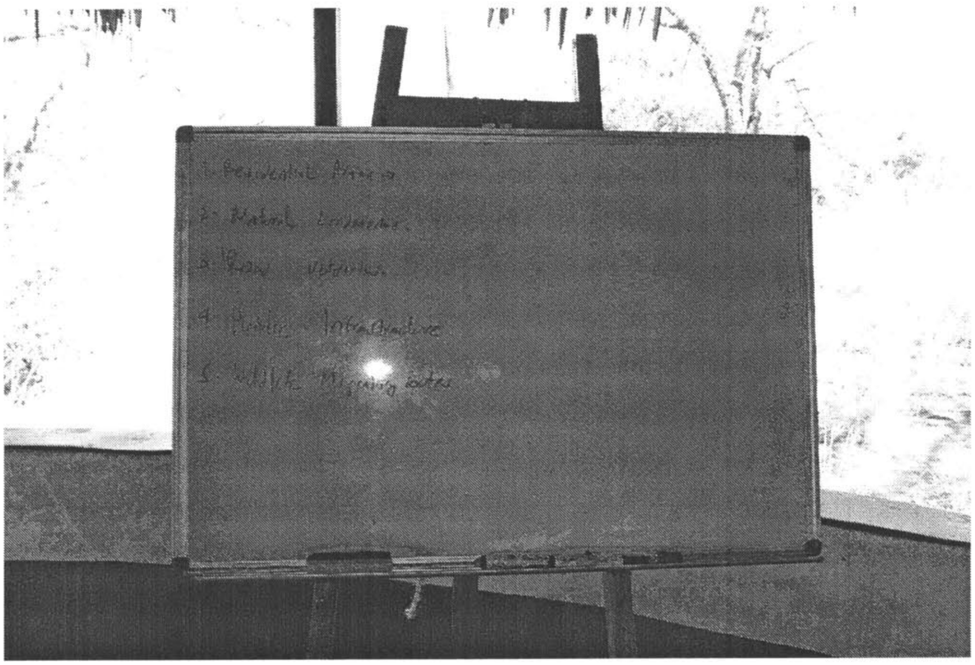
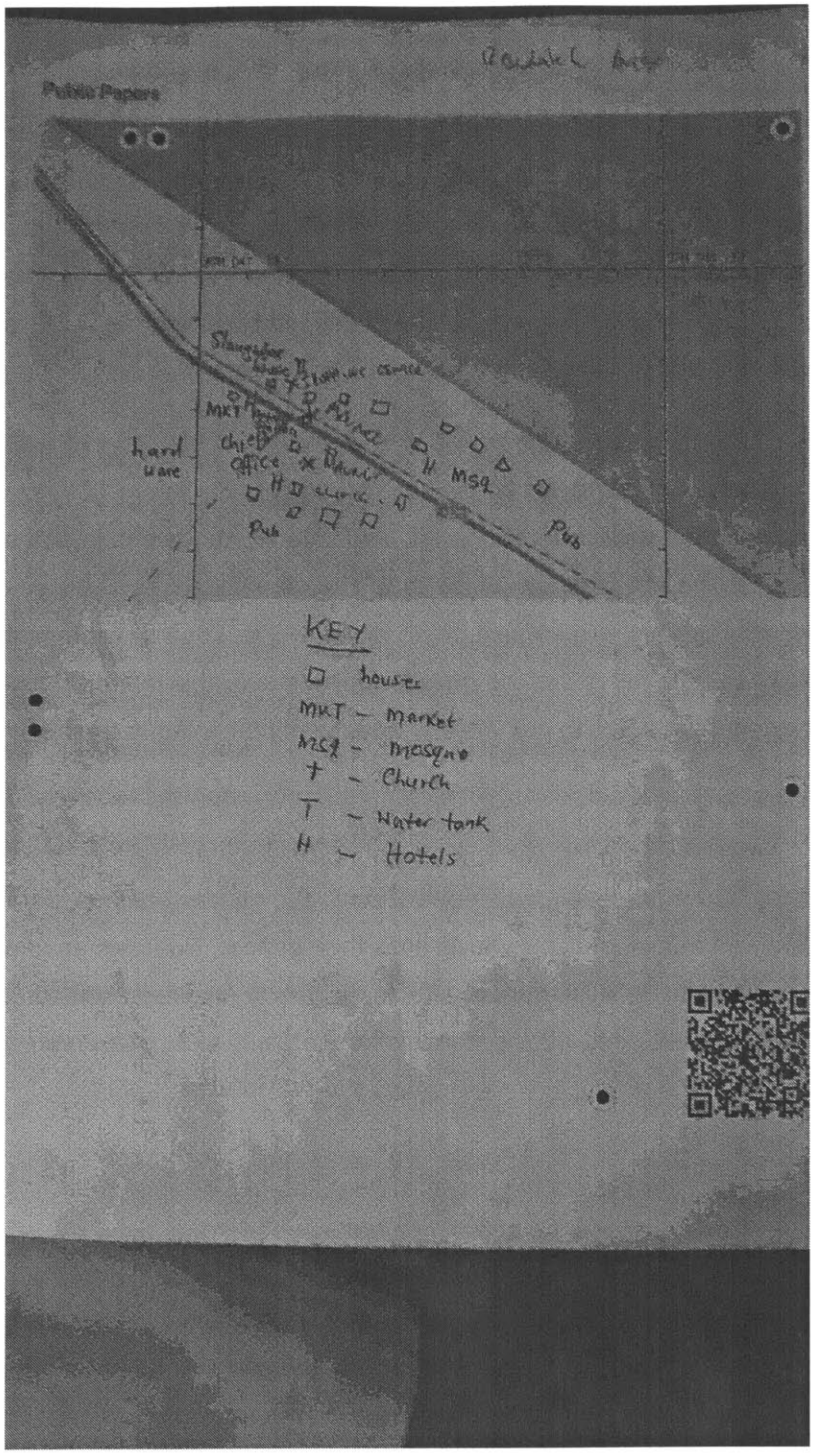


Image showing map categories based on importance

8.3.2 Mapping Tasks

After identifying areas some of these communities were interested in mapping, members were provided with mobile handsets with the Mapjack application installed and asked to head out to different areas and collect data. Given some participants did not have access to mobile phones, paper based maps were provided to some. The image below depicts a paper based map from one of the participants from the Tsavo group.



Public Papers
hard ware
MKT
chief office
clinic
H
MSQ
Pub
Pub
KEY
☐ houses
MKT - market
MSQ - mosque
+ - Church
T - Water tank
H - Hotels

Field Mapping by a Tsavo Group Participant

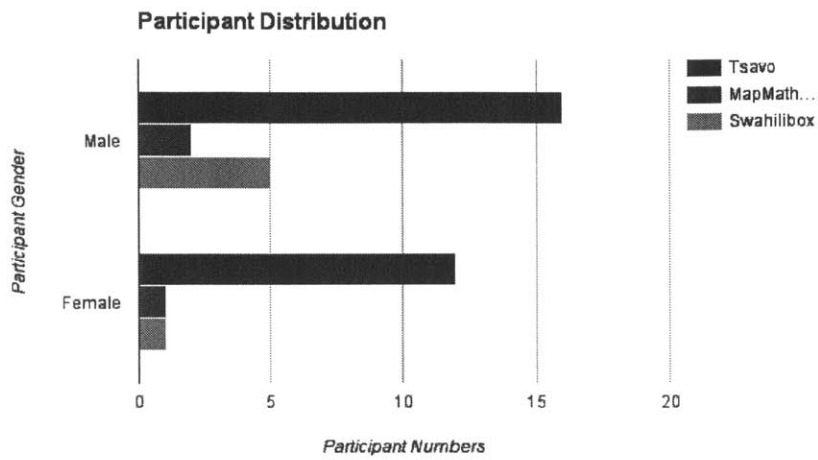
Participants were required to walk around and compare the actual map they had and make decisions on what is not represented or should be represented. In the case of those who had mobile phones, they were required to add the points using the custom mobile application. Before heading out to map, a quick demo of the mobile application was made to all participants. Participants were asked to compare what they see physically and if they did not see it on the map, add it by double clicking on the map. If they noticed something was misrepresented, they were required to add a bug and write a comment.This process took on an average one hour.

8.3.3 Study Questionnaire

After carrying out data collection in the field, interviews were held with participants to get their perspective on the effectiveness of the tool. The questionnaire contained questions pertaining to spatial literacy, easy of use of the tool and questions on representation of geographic features. The full text of the questionnaire is available in the appendix. Subjects were also provided the chance to ask me any questions pertaining to the study.

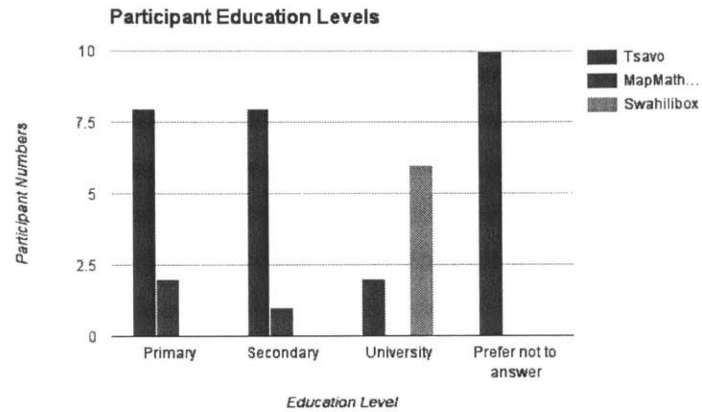
8.3.4 Results

The study had more male participants though the gender balance is commendable. The graph below shows participant numbers vis a vis their gender. Reasons as to why gender is also important in this study is because gender affects matters spatial representation: what might matter to male participants may not necessarily be the same to female participants.



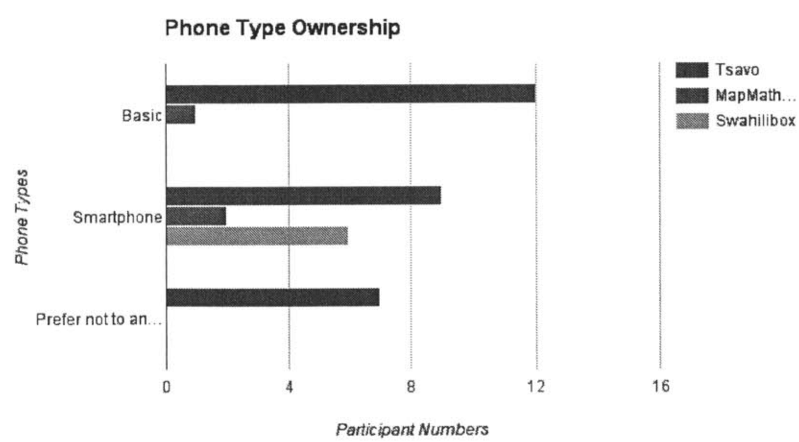
Participant distribution

Another important aspect assessed during the study is the education level of participants. Does education have a role to play in how people understand and read maps? The graph below shows education levels by participants. The Tsavo group is the only group with varying ranges of educational skills. All in MapMathare had undergone some form of Diploma while those surveyed at Swahilibox were mainly undergraduate students. Several members from the Tsavo community opted not to answer the question as education is perceived as a skill that is very valuable to have and for those who did not go to school there is an element of personal guilt or shame.

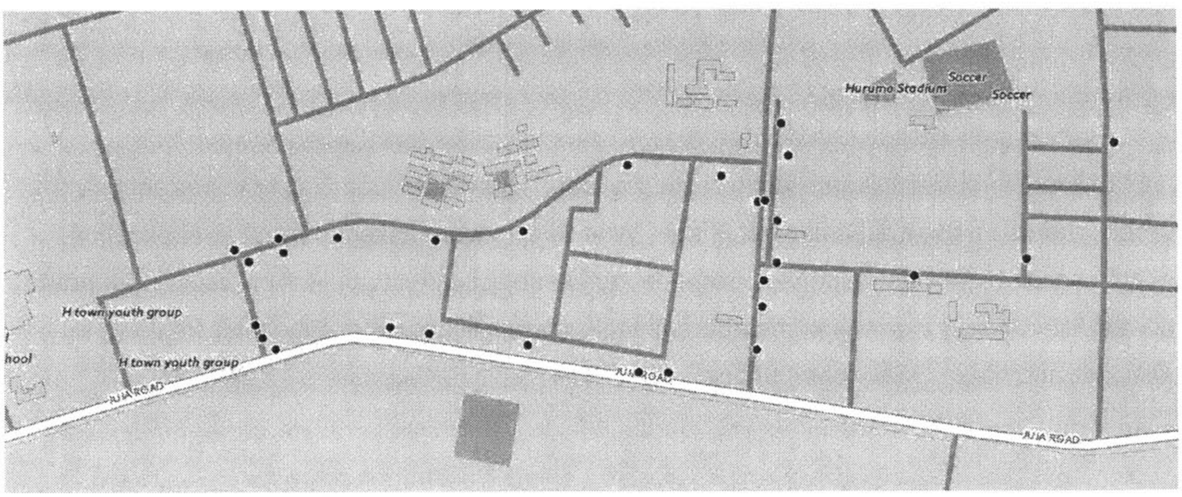


Subject Education Levels

To evaluate the success of such an application in the hands of citizens in emerging economies, subjects were asked about the access to devices they have. This is to understand the kinds of devices participants have access to. The image below shows the distribution of devices participants had access to. It seems from the data that members who are located within urban areas are likely to have smartphones as opposed to those located in rural areas. All members of Swahilibox and two members of MapMathare had smartphones.

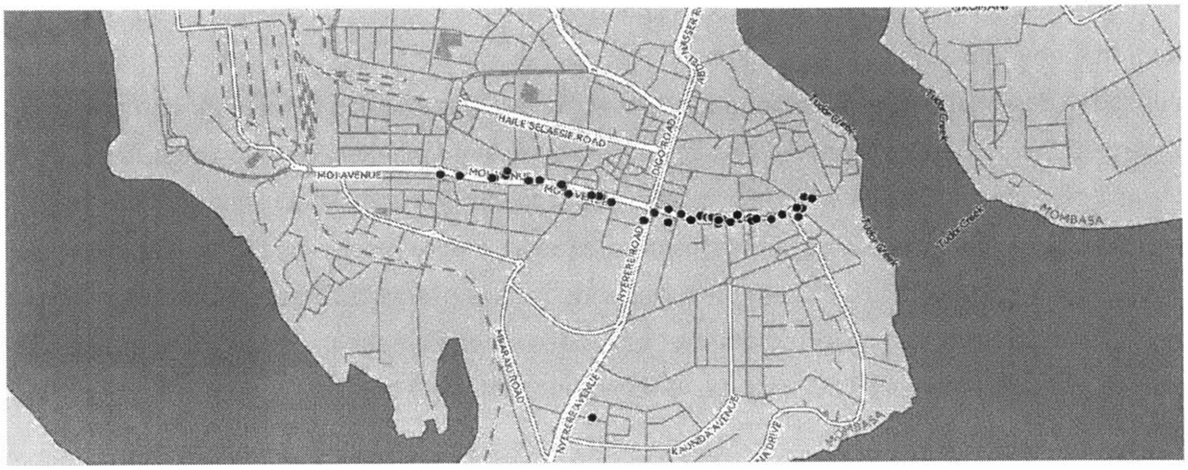


After data was collected by members of the MapMathare community group, the data was digitised and the map below shows edits made. Edits are color coded according to participant i.e. those in red were collected by one participant while those in green by another. Sadly the data from one of the participants could not be extracted from the mobile phone hence only two color codes depicted below. Of importance is how many points it took one member to map as opposed to the other participant. All participants were given 45 minutes to map. The data collected also varies. Of interest is how different participants consider what is important to map. One of the participants, depicted in green, collected more mobile money locations than the other. The subject secondly collected public infrastructure and local businesses to the map. The other subject coded in red collected more religious and entertainment amenities more than the other. This shows how subjective maps and representation could be.



MapMathare Contributions by the three data collectors

Contributions made by members of SwahiliBox were widely varied. This also proved to be a great test case as it is the only data set that registered an error. A locational error can be seen at the bottom of the image and this could be attributed to GPS locking issues. Of interest is the tendency by participants from SwahiliBox to collect entertainment related points of interest. Financial related points of interest were also frequently collected by the different participants. One of the participants was interested in editing street names, and could not as the application only allowed for points based editing.



Contributions by SwahiliBox Members

8.3.5 Tool Evaluation

Participants were asked on their experiences using the tool. Simon Kokoyo from MapMathare expressed his satisfaction with the tool. "This offline map is very important as now I can see what is on the map and what is not." When users were asked to rate the application, the table below highlights responses from the study subjects. Most application users from both the MapMathare and Swahilibox members felt that the application made it easier for them to participate and make OSM contributions.

	MapMathare	SwahiliBox
Participant 1	10	8
Participant 2	8	8
Participant 3	10	8
Participant 4		8
Participant 5		7
Participant 6		9
Average	9.333333333	8

On a scale of 1-10 Ratings on using Mapjack to Make OSM Contributions

When participants were asked if offline mapping was important to their consumption and contribution of mapping services, most strongly agreed that the provision of spatial context when collecting data was very useful. When asked if they had used or come across tools that provide them offline rendering for spatial context when collecting data, all participants in the study mentioned they had never come across any tools. It is important to to note, that for those in the SwahiliBox cluster who had experience with Google Maps mentioned that they had never accessed any tools to make edits to Google Maps or consume Google Maps services in an offline environment. The table below highlights the rating members of MapMathare and Swahilibox placed on spatial context when collecting data. Tsavo members also answered this question, but most used offline paper based maps rather than the Mapjack mobile application.

	MapMathare	SwahiliBox
Participant 1	9	7
Participant 2	8	8
Participant 3	9	9
Participant 4		
Participant 5		7
Participant 6		9
Average	8.666666667	8

Rating on the importance of Spatial Context when collecting data

Participants were asked on some of the challenges they faced using the tool and one mentioned, "The tool does not provide access to imagery, if one could also get imagery, it would be more useful especially in areas where there are very few contributions". Another mentioned, "You need to allow for taking photos and enable users to add locally generated landmarks on the offline map."

When participants were asked to rank how well they think they are spatially represented, many had varying opinions on their level of spatial representation and identity. The table below shows ratings on their spatial representation. A couple of points could be derived from this table. MapMathare has been in the process of updating OSM data over a period of the past two years. The involvement of some of the participants in this study point to a state where most are of the opinion that they are now well spatially represented with the exception of one. This is also evident in the edits as depicted in the MapMathare map highlighted earlier. SwahiliBox participants have a general sense of under representation. This is evident in the data. For all the participants from SwahiliBox, none had ever edited OSM before. Some of the feedback provided was that not only is there lack of sufficient data, but also elements of misrepresentation abound.

	MapMathare	SwahiliBox
Participant 1	9	5
Participant 2	5	9
Participant 3	9	4
Participant 4		1
Participant 5		6
Participant 6		5
Average	7.666666667	6

8.4 Results Analysis

8.4.1 Spatial Literacy

The range of users who tested the tool was wide and provided very valuable feedback on the tool. For instance, members of the Tsavo community consisted of older participants with very little formal qualifications. On average it took 10 minutes for those in this community to understand the map. This is unlike the MapMathare and Swahilibox members who took generally took less than two minutes to understand the map. Based on feedback of participants interviewed by the Tsavo group, those who undertook geography classes seemed to have a better understanding reading and interpreting the map. Luckily, since the Tsavo group performed the data collection in groups of four, those who had not formally studied geography had an opportunity to learn from those who had a better understanding. From the tests performed by all the groups, it seems education in Geography plays an important role in determining how people understand and interpret maps. This sample may not be the best to answer this question but perhaps a larger study with subjects who study in different educational systems and age groups would help better answer this question. For instance, one could conduct a study of subjects who undertake Nigeria's educational system versus the Kenyan education system where Geography is mandatory for both elementary and high school.

8.4.2 Misrepresentation

Part of the process by participants was adding to the map items that they considered forms of misrepresentation. They simply did this by clicking on the map and registering a bug and saving the details. When approaching this research, I considered misrepresentation as in a large part a function of symbology. Misrepresentation includes having a shrine represented as a church or a mobile money agent being represented as an ATM, for example. During the deployment, on inspecting data from members of the Swahilibox and MapMathare communities, other forms of misrepresentation come to the fore. One example is in the use of labeling and feature types. For instance, the name of a street Al' Nur was represented as "alnur" (on the OSM map). The street was named after a local politician and in written forms of communication his name was spelled as Al' Nur. This is likely to result on how people search for data and also the element of importance to labels provided on maps. The other type of misrepresentation highlighted was a market that was placed on the map as Kongowea. To many, Kongowea is a large physical space and the best way to represent it is in the form of a polygon rather than a point as it was on the map. This highlights the level of importance such users place on representation. For one of the subjects, representing the market as a point rather than an area was grossly misrepresenting it given the importance of such markets to such communities.

8.4.3 Offline Mapping

All respondents interviewed expressed pleasure with offline mapping. When respondents were asked if offline map rendering and the provision of spatial context helped them collect more data, all mentioned that offline mapping as implemented on Mapjack helped them discover what is mapped and what was not. "The fact that I can now see what is on the map and what is not as I walk around collecting data is very helpful", said one of the respondents. "This is very useful especially since most of the data we did three years ago is now outdated and this offline map helps us see what needs to be updated", another member of MapMathare reported. Another highlight while working with the members of the MapMathare group is when they expressed interest in having the application installed on their local phones for them to continue using it. They also provided some suggestions on how to improve the data collection by changing the tool to support more than one feature type data collection. The ability to actually pinpoint on the map and trace either a point, line or polygon was a major feature request by most interviewed. Another potential improvement was the lack of querying metadata. For the experienced OSM contributors from MapMathare, Mapjack needs to provide mechanisms to query the metadata in an offline environment.

9 Future Work

Improvements such as support for multiple feature types, better metadata editing and better categorization of the data collected was suggested by members of the different communities used to test Mapjack. One major improvement is the provision of support for multiple feature types: points, line and polygons. During the test only the collection of points was supported. More needs to be done to provide a simple interface to allow users to trace features on the mobile phone. Another improvement suggested by the users was broadening the categories while filling in metadata tags. Perhaps using icons could make this process easier when entering metadata. Alternatively, another approach could be introducing nested categories which provide more metadata options as one traversed the different levels, moving from defined models to loosely coupled models. This could involve providing an open, editable schema on the mobile phone and a toolkit to write and read this custom schema.

Ensuring that locally generated landmarks or points of interest can easily be searched and also symbolized on the map locally is also another area where more work needs to be done. This was based on feedback by one of participants that they may have their own entities which they want to map locally and also want to have them represented to them as they see fit. For instance, someone collecting Mosque locations may not like the default symbology that is provided with the offline map. They could have the ability to define their own symbology and have it represented as they deem fit on the map. This would involve a lot of work and indeed goes to the heart of the question of misrepresentation: who gets to determine which icons represent which features on a map? I think this feature request is kind of a mechanism to deal with matters of iconic misrepresentation.

Addition of multimedia content was also identified as a possible improvement. Being able to take photos or videos and attach them to specific feature sets is seen as a good mechanism to crowdsource street level images. Equally important was a request where one of the participants mentioned that providing a framework where maps could be personalised and easily shared could improve the tool and also improve contributions. This means one would have the freedom to determine how they want to consume cartographic content, based on what they view as important. A mechanism to help people consume cartographic content based on their frequent needs was also touted as a mechanism to improve local grassroots mapping.

How metadata is handled may need to change. This was based on the request by some of the users to query metadata offline. This proved to be a challenge due to the compressed file formats in use. A proposed mechanism to deal with this is to separate the feature set data from the metadata but have them both indexed. Rather than having a compressed OSM format file containing both metadata and the feature sets, having an indexing database referencing a feature set entry to a metadata entry could make such querying easier. I need more time to implement this and more tests need to be performed on the benefits that such an architecture would bring forth.

More studies need to be carried out upon making some of the changes suggested. Another deployment will be carried out with some changes to maintain momentum and to gain more insights on how such civic actors see in the importance of maps. More need to be done on participatory divides and perhaps setting up an OSM atlas of digital inequality could provide more insights on other possible divides not highlighted in this thesis.

10 Conclusion

This thesis shows that new forms of the digital divide are affecting how people create and participate in creating cartographic content. Mapjack is a Android based application that enables citizen cartographers to create, share and disseminate cartographic information through the use of a mobile device. It employes offline rendering and editing techniques to bring about more engagement with the wider OpenStreetMap community. By using offline rendering techniques, the cost of access to cartographic data is lowered. Though the mobile phone alone is not sufficient to address most of the underlying issues facing OpenStreetMap contributions, it has the greatest potential given its prevalence in both emerging and already developed economies. Mapjack is a tool that can be used to meet the needs of different groups. As evidenced in the three test groups, none of the three groups have similar mandates but the tool was deemed useful to all the three. These needs could vary from data collection for conservation to data collection for civic action.

Mapjack has the potential to provide more motivation and participation to people who would not ordinarily consider themselves citizen cartographers. With more resources to develop some of the enhancements and changes requested by some of these community members, this could be realized as a complete open source tool whose usage would traverse different geographic boundaries and data collection limitations.

11 Bibliography

Brown, G., Kelly, M., Whitall, D., 2014. Which “public”? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. Journal of Environmental Planning and Management 57, 190–214. doi:10.1080/09640568.2012.741045

Communication Commission of Kenya, 2013. Quarterly Statistics Report. URL. http://216.154.209.114/resc/downloads/Q4_201213_STATISTICS_final_25th_oct_2013.pdf

Graeff, Erhardt Charles. “Action Path: A Location-Based Tool for Civic Reflection and Engagement.” Massachusetts Institute of Technology, 2014. *Google Scholar*. Web. 7 May 2015.

Flanagin, A.J., Metzger, M.J., 2008. The credibility of volunteered geographic information. GeoJournal 72, 137–148. doi:10.1007/s10708-008-9188-y

Fitcher, D., 2009. What is a Mashup?, in: Library Mashups : Exploring New Ways to Deliver Library Data. Information Today, Inc.

Goodchild, M.F., 2007a. Citizens as sensors: the world of volunteered geography. GeoJournal 69, 211–221. doi:10.1007/s10708-007-9111-y

Goodchild, M.F., 2007b. in the World of Web 2.0. International Journal 2, 24–32.

GSMA, 2012. Mobile telephony and taxation in Kenya

Habermas, J., 1989. The structural transformation of the public sphere: an inquiry into a category of bourgeois society. MIT Press, Cambridge, Mass.

Hargittai, E., 2008. The Digital Reproduction on Inequality, in: Grusky, D. (Ed.), Social Stratification: Class, Race, and Gender in Sociological Perspective. Westview Press, pp. 936–944.

Sieber, R., 2006. Public participation geographic information systems: A literature review and framework. Annals of the Association of American Geographers 96, 491–507.

Tulloch, D.L., 2007. Many, many maps: Empowerment and online participatory mapping. First Monday 12.

Tulloch, D.L., 2008. Is VGI participation? From vernal pools to video games. GeoJournal 72, 161–171. doi:10.1007/s10708-008-9185-1

"10 Myths About M-PESA: 2014 Update." *CGAP*. N.p., n.d. Web. 7 May 2015.

"Fieldpapers" *Stamen*. N.p., n.d. Web. 7 May 2015.

Fogg, Brian J. "Persuasive Technology: Using Computers to Change What We Think and Do." *Ubiquity* 2002.December (2002): 5. Print.

12 Interview Questions

Research on New Forms of The Digital Divide Affecting Cartographic Production

Subject Code:

Mobile Phone Number: Education Level:

1. Do you have a mobile phone?

2. What is the make of the mobile phone?

3. Do you have any prior experience in contributing to OSM?

4. What tools have you do you use mainly when editing the OSM basemap?

5. What motivates you to map your local community?

6. How frequently do you map your local community?

7. Do you use your phone to make contributions to OSM and which mobile tools do you use?

8. For how long have you been a data collector?

9. How would you evaluate Mapjack on a scale to one to ten in making OSM contributions?

10. Does the ability of Mapjack to provide spatial context and offline map rendering improve or help you collect data better in your opinion? How better on a scale of 1-10? Do any of the tools you use provide offline rendering?

11. Does the current offline map have a generally fair representation of your area on a scale on 1- 10? Do you consider the symbology a great form of representation of your area?

12. What important thing do you think is missing that should be on the map?

13. Did you receive any notifications while using the tool?

14. Did you act upon any of the notifications?

15. Any improvements you think should be made to the tool?