Guidelines for creating framework data for GIS analysis in low- and middle-income countries

Prestige Tatenda Makanga
Department of Geography, Simon Fraser University
Department of Surveying and Geomatics, Midlands State University

Nadine Schuurman
Department of Geography, Simon Fraser University

Charfudin Sacoor
Centro de Investigação em Saúde de Manhiça

Helena Boene
Centro de Investigação em Saúde de Manhiça

Peter von Dadelszen
Department of Obstetrics and Gynaecology, St George's, University of London

Tabassum Firoz
Department of Medicine, University of British Columbia

Key Messages

- While some framework data are available in LMICs, there is a lack of coordinated effort to create and share these data to support health GIS research in these settings.
- Manual digitizing can be used to generate framework data, and is increasingly becoming cheaper and faster due to widely available free satellite imagery and open mapping standards that allow distributed data capture.
- Efforts to create framework data should be done in collaboration with local mapping authorities that have the mandate for creating these data at scale.

Health sciences research is increasingly incorporating geographic methods and spatial data. Accessing framework data is an essential pre-requisite for conducting health-related geographic information systems (GIS) research. However, in low- and middle-income countries (LMICs) these data are not readily available—and there is a lack of coordinated data creation and sharing. This paper describes a simple set of strategies for creating high-resolution framework data in LMICs, based on lessons from a maternal health GIS project—"Mapping Outcomes for Mothers"—conducted in southern Mozambique. Data gathering involved an extensive search through public online data warehouses and mapping agencies. Freely available satellite image services were used to create road centrelines, while GPS coordinates of households in the study area were used to create community boundaries. Our experience from this work shows that manual digitizing is becoming cheaper and faster, due to increased availability of free satellite image services and open mapping standards that allow for distributed data

Correspondence to/Adresse de correspondance: Prestige Tatenda Makanga, Department of Geography, Simon Fraser University, 8888 University Drive Burnaby, BC, V5A1S6. Email/Courriel: pmakanga@sfu.ca

capture. Involving mapping agencies in data capture processes will likely promote the scaling up of framework data creation in LMICs. This will benefit health GIS research in these settings.

Keywords: framework data, health GIS, data capture, global maternal health, manual digitizing

Lignes directrices pour créer des données de référence pour les analyses de SIG dans les pays à revenu faible ou intermédiaire

Un nombre croissant de travaux de recherche en sciences de la santé reposent sur des méthodes géographiques et des données spatiales. L'accès aux données de référence est une condition essentielle pour réaliser des recherches en santé utilisant un système d'information géographique (SIG). Cependant, dans les pays à revenu faible ou intermédiaire (PRFI), ces données sont peu ou pas disponibles et aucune coordination n'existe pour les concevoir et les partager. L'objet de cet article est de présenter des stratégies simples pour concevoir des données de référence de haute résolution dans les PRFI à partir des enseignements tirés d'un projet de SIG portant sur la santé maternelle, « la cartographie des retombées pour les mères » dans le sud du Mozambique. La collecte de données partait d'une recherche poussée menée dans les archives de données publiques en ligne et des agences de cartographie. Les services d'imagerie par satellite en libre accès ont été utilisés afin de concevoir des lignes médianes d'axes routiers, tandis que les coordonnées de GPS des ménages de la zone d'étude ont servi à créer des frontières communautaires. Il ressort de ce travail que la numérisation manuelle de données est devenue plus économique et plus rapide, en raison de la disponibilité accrue de services d'imagerie par satellite gratuits et des normes ouvertes de cartographie qui permettent la saisie de données réparties. L'implication des agences de cartographie dans les processus de saisie de données devrait contribuer au renforcement de la conception de données de référence dans les PRFI. Ceci profitera aux recherches en santé utilisant un SIG dans ces milieux.

Mots clés : données de référence, SIG de la santé, saisie de données, santé maternelle mondiale, numérisation manuelle

Introduction

Framework data for geographic information systems (GIS) analysis have been defined as "geospatial data themes identified as the foundation upon which all other data layers are structured and integrated for analysis and application" (Berendsen et al. 2010, 10). In most high-income settings, maps and framework GIS data at large scales are readily available. However, for much of the world, these data are largely unavailable, and those data that are available are often stored away in private databases of non-governmental organizations and corporations (Von Hagen 2007). In low- and middle-income countries (LMICs), particularly in Africa, the lack of good framework GIS data, as well as the cost associated with creating them, have been acknowledged as the major limitations to the use of GIS in health research (Aimone et al. 2013; Kim et al. 2016). There have been some efforts to fill these data gaps through programs like the Global Spatial Data Infrastructure (GSDI) initiative (Stevens et al. 2005), Mapping Africa for Africa (Gyamfi-Aidoo et al. 2005), and the United Nations Economic Commission for Africa (UN ECA and DISD 2001). Despite all these efforts there remains a glaring need for framework GIS data, as well as new protocols that aid the development and sharing of these data.

There is a rapid growth in the use of geographic thought and spatial analysis techniques within the health sciences (Richardson et al. 2013). Health research typically requires framework GIS data as the basis for overlaying other datasets such as those for health facilities and population-level disease prevalence (Tanser and Le Sueur 2002). Without such data it is difficult to communicate population-level health problems and identify areas with the greatest need for health interventions. Thus, spatial epidemiology is founded on the framework data for the area under investigation. Typically, framework data come from mapping agencies rather than existing health research (UN ECA and DISD 2001; Rajabifard et al. 2006). However, in some instances where framework data do not exist, health researchers have had to create their own (e.g., Bailey et al. 2011). In instances where village health workers access their communities on a regular basis, they

can be used to map neighbourhoods in a manner that is cheaper than hiring professional GIS practitioners to do the same (Munyaneza et al. 2014).

Due to the general unavailability of framework data in LMICs, health researchers have sometimes resorted to using more readily and freely available coarse-resolution data to model the geography of health-related issues. For example, 90 m digital elevation models have been used together with land cover data to model access to care using friction surfaces (Masters et al. 2013; Fogliati et al. 2015). Access to health care services has also been modelled by Friedman et al. (2013) "as the crow flies," through the creation of buffer zones around health facilities to avoid the process of creating road data required for more precise estimates of travel time. These approaches, while useful at very small scales, produce results that are simply not useful for understanding health trends at the community level. Semi-automated methods such as feature extraction from satellite images have also been explored (Awad 2013), but the costs associated with post-processing and acquiring the high-resolution satellite images remains a prohibiting factor.

Despite the free availability of coarse-resolution GIS data from public sources, detailed framework data remain essential for modelling geographic patterns in health (Schuurman et al. 2006), particularly for facilitating targeted community health programming in highly burdened settings that have limited resources. This article introduces a series of strategies for creating framework GIS data in a data-poor setting based on the experiences from the Mapping Outcomes for Mothers (MOM) project in a largely rural region of southern Mozambique. These data creation strategies are an important contribution to the execution of health-related GIS research, spatial mapping, and analysis in LMIC settings where framework data are not readily available.

Project context and GIS data needs

The MOM project was set in 12 administrative regions in the southern part of Mozambique: four in Maputo province and eight in Gaza province. The overall aim of MOM was to explore the community-specific factors that elevate risk in pregnant women, resulting in maternal deaths or instances of severe maternal morbidity. We also endeavoured to identify community-specific factors that promote healthy pregnancy

outcomes. Part of this process included modelling access to maternal health services—as well as how this access was affected by the seasonal floods and wet weather that plague this study region almost every year. MOM was undertaken in partnership with the Community Level Interventions for Pre-eclampsia (CLIP) cluster randomized control trial (Clinical-Trials.gov ID: NCT01911494), which at the time of writing this paper was evaluating a care package delivered by community health workers in an effort to reduce all-cause maternal mortality and severe morbidity in Pakistan, India, Nigeria, and Mozambique. Two key datasets were created as part of the MOM work, including a detailed set of community-level roads for modelling spatial access to maternal care services, and a set of high-resolution community boundaries for quantifying community-specific risk and resilience factors related to maternal outcomes.

Modelling spatial access to maternal care

Modelling potential geographical access to health services is the main application of GIS in maternal health research (Ebener et al. 2015; Makanga et al. 2016). A detailed road network dataset is normally required for this work—in order to trace the actual paths that are potentially used to navigate through space. However, alternate methods for quantifying spatial access to the closest facility that require less detailed data can be used in the absence of good framework road data. Examples include calculating fixed distance buffers around health facilities (Ivers et al. 2008), and creating friction surfaces based on low-resolution and freely available digital elevation models and land use data (Masters et al. 2013). These methods have all been shown to produce comparably accurate results for identifying the closest facilities (Nesbitt et al. 2014). However, it has also been demonstrated that patients do not necessarily access their closest facility (Alford-Teaster et al. 2016), and that there are many other pervasive factors that influence which facility will be used in the time of need. Moreover, patients do not necessarily walk or drive as the crow flies.

The MOM study sought to extend current models of access to care by accounting for the hierarchical nature of travel through the health referral chain (i.e., primary health facilities to secondary health facilities to tertiary health facilities) instead of simply identifying the closest facilities. Modelling potential spatial access using this hierarchical manner more

closely matches the reality of women's travel through the health care system and required a more detailed inspection of the pathways of travel used when seeking maternal care—hence our need for a high-resolution network of community-level roads. We also sought to factor into our analysis the impact of precipitation and floods on access to maternal care. Therefore, we needed a dataset that described the actual roads (and their condition) that women would travel on (e.g., paved/unpaved) to quantify how the road infrastructure would be affected by weather conditions, and also to identify the specific community road segments that would not be usable in the event of flooding.

Modelling community-level risk and resilience in maternal health

The use of geographically explicit statistical techniques is increasingly being recognized in health GIS research as this approach adds value by illustrating how associations with disease patterns change across space (e.g., Aguilera et al. 2007; Shoff et al. 2012; Owoo and Lambon-Quayefio 2013). These kinds of analyses require data to be aggregated into small geographical units prior to analysis in order to reveal the underlying spatial patterns that are normally masked into single values (e.g., beta coefficients and R^2 values) (Makanga et al. 2016). An example of one of these techniques, which was used in the MOM project, is geographically weighted regression (GWR). Unlike ordinary least squares regression, GWR evaluates the non-stationarity of parameter estimates across space (Shoff et al. 2012). GWR typically requires many data points (neighbourhoods) for the valid estimation of the changing parameter values (Fotheringham et al. 2001).

The MOM study used GWR to explore the spatial epidemiology of maternal ill health in the study area. This entailed elucidating the associations between possible risk and health promotion factors as measured through the baseline study of the CLIP trial, as well as how these associations changed over different communities. A host of socio-cultural variables including financial support, emergency transport availability, and financial decision making were collected for every household in the study area, with the intention of aggregating these to community-level scores. There was, therefore, a need to create high-resolution community boundaries as a basis for doing further geo-statistical analysis and identifying the place-specific patterns relating variables to rates of maternal deaths.

Other datasets

Two other datasets were required for this work, including a spatio-temporal dataset for precipitation spanning a retrospective 1-year period starting from the point when the CLIP baseline survey was conducted. We also required data on the flood extent for the same timeframe. These data helped in modelling the seasonal impact of weather elements on access to maternal care services in the study area.

Data sources and data creation

Existing data sources

The first step for data acquisition was searching through public databases that host spatial data, such as DIVA-GIS and OpenStreetMap, and identifying relevant datasets. We also met with personnel from the Mozambique National Cartography and Remote Sensing Centre (CENACARTA), and other local mapping institutions who shared their available data. The data were assessed to evaluate if they needed re-formatting for our needs. Through this process, we were able to quantify the data gaps that needed to be filled through alternate methods. A key lesson from this data-building process was the importance of liaising with local data stewards as well as drawing on open geodata repositories.

A summary of the freely available public datasets acquired for this project is provided in Table 1. There were multiple versions of the same data, none of which had comprehensive metadata to describe how they could be used. Both the street data and administrative boundaries were only available at very small scales; highways and paved main roads were the best roads available, while the best available administrative boundaries were at the administrative post level. Lower level boundaries for localities and neighbourhoods were not available at any of the public data sources or CENACARTA.

Capturing road data

Gaps in the road datasets were filled through a process of manual digitization. We set up a

Table 1
Datasets acquired from public databases and CENACART

| Data Source	Datasets Acquired
Public Databases (e.g., DIVA-GIS, OpenStreetMap, UN FAO clearinghouse, CENACARTA website)	Street data • Highways Administrative boundaries • National, provincial, district, and administrative post
OpenStreetMap	Street data • Highways, major roads, and some minor roads
CENACARTA (National Mapping Agency)	Street data • Highways, major roads Administrative boundaries • National, provincial, district, and administrative post
Bureau of Statistics	None
Health Ministry	Health facility coordinates
CLIP Project	Households • GPS coordinates • Household IDs
Manhica Research Centre (Local research partner)	List of all neighbourhood IDs linked to neighbourhood names, locality, and administrative post Street data • Highways, major roads • Administrative boundaries • National, provincial, district, and administrative post
Global Flood Observatory	Daily flood extents for the study area (for quantifying the impact of flooding on access to health facilities)
Famine Early Warning Systems Network (FEWSNET)	Daily precipitation estimates for the study area (for quantifying the seasonal impact of precipitation on access to health facilities)

custom data capture platform (Figure 1) based on the ArcGIS suite of software, that met our data capture needs and also allowed for data capturers to use a familiar software platform. The idea of a custom interface for data capture was also to demonstrate that the process of setting up custom applications has become much quicker, and that researchers who require data capture tools that are not part of the design in existing free platforms (e.g., OpenStreetMap) have an option to make their own.

An ArcSDE multiuser geodatabase was designed to allow for multiple remote users to digitize the data concurrently and sync the changes centrally to the MOM server in near real time. A freely available satellite image service from Bing Maps was used as reference for tracing out the new road features. The images for the study area were last updated in April of 2012 and were available at 60 cm resolution at the 18th zoom level (Bing Maps 2016; MSDN 2016). A Web Feature Service (WFS) was used to render the contents of the geodatabase to the multiple users for editing through the ArcGIS server platform. The WFS is an open standard for rendering and manipulating

geographic vector features through the web and independent clients (Strobel et al., 2016).

There were two possible ways of accessing the WFS: (1) through secure access from a remote ArcGIS desktop client, or (2) through our custom browser-based MOM application, designed using the ArcGIS viewer for flex (Figure 2). In the first instance, the data capturer would set up a local ArcSDE geodatabase that would allow for data to be downloaded onto their computers from the MOM server, for editing. Upon completion of a digitizing session, data would be synced back to the MOM server. In the second instance, users would access the WFS and do all edits through a web browser.

Twenty-six student volunteers were initially recruited to help with the digitizing. However, we realized early into the data capture process that there were a lot of quality control challenges. We thus chose to work with only four students who were compensated for their time. Estimations of time spent digitizing were based on timestamps attached to each digitized feature. All students were given basic training to minimize data capture errors and to ensure uniform interpretation of imagery
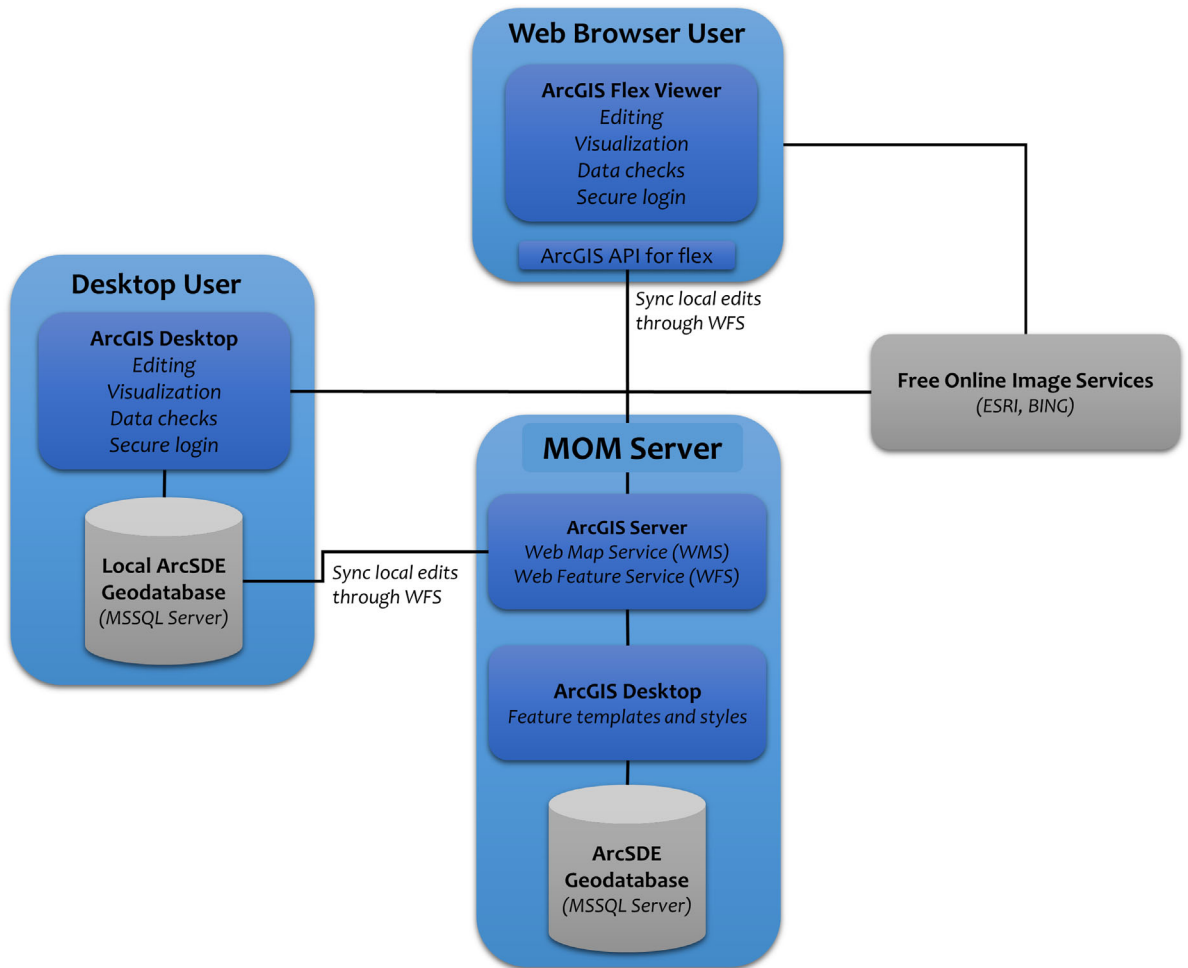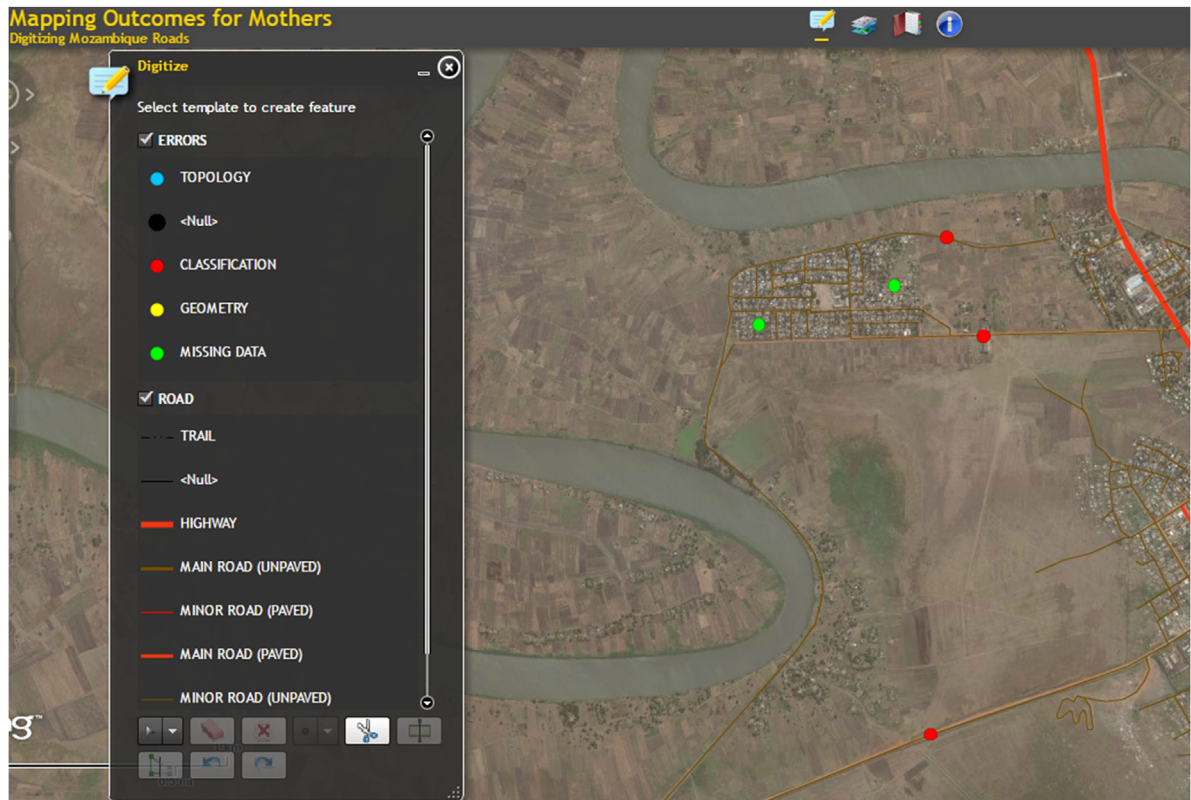
Figure 1
Architecture of the data capture platform.

during the data capture process. The data capturers also checked each other's work for geometric, topological, and classification errors, as well as omitted roads. One of the data capturers was solely dedicated to checking all the digitized features for these errors. Two staff from CENACARTA also volunteered to check the data. Classification of road types was done by consensus between all participating parties based on local knowledge and interpretation of the imagery.

A total of 15,014.4 km of road length was manually digitized and checked for the data capture

errors described earlier (Table 2). The roads were classified into six themes: highways, paved main roads, unpaved main roads, paved minor roads, unpaved minor roads, and trails—with most roads (71.2%, 10,694.7 km) being unpaved minor roads. The total time for the digitizing process was 179 hours, which translates to roughly 22 8-hour days. This is significantly less time than what we expected for such high-resolution data capture.

Although OpenStreetMap data have been used in other studies in which these data were confirmed as an accurate representation of what's on the ground

Figure 2
Flex-based web viewer for digitizing roads data.

(Ferguson et al. 2016), this was not the case for our study area, as illustrated by the huge data gaps in Figure 3. At the time of writing this paper, arrangements are underway to publish all the new data to the OpenStreetMap platform to serve as a contribution to a wider GIS audience needing access to data in the study area.

Table 2
Summary of new roads data

Road classification	Total length (km)	% of all roads
Highway	788.3	5.3
Main road (paved)	610.1	4.1
Main road (unpaved)	1978.4	13.2
Minor road (paved)	479.5	3.2
Minor road (unpaved)	10694.7	71.2
Trail	463.4	3.1
Total length (all roads)	15014.4	
Total time for digitizing (hr)	179.0	

Capturing community boundaries

Community boundaries were created from GPS coordinates of households in the study area that were acquired from the CLIP baseline survey. Each household in the study was assigned a unique 10-character household identification (ID) that indicated the administrative post, locality, neighbourhood, and household number (Figure 4). A neighbourhood is the smallest administrative unit and the administrative post is the largest administrative unit for this study. Multiple neighbourhoods are contained in a single locality, and multiple localities in an administrative post. Ethics approval for this study only allowed us to access information
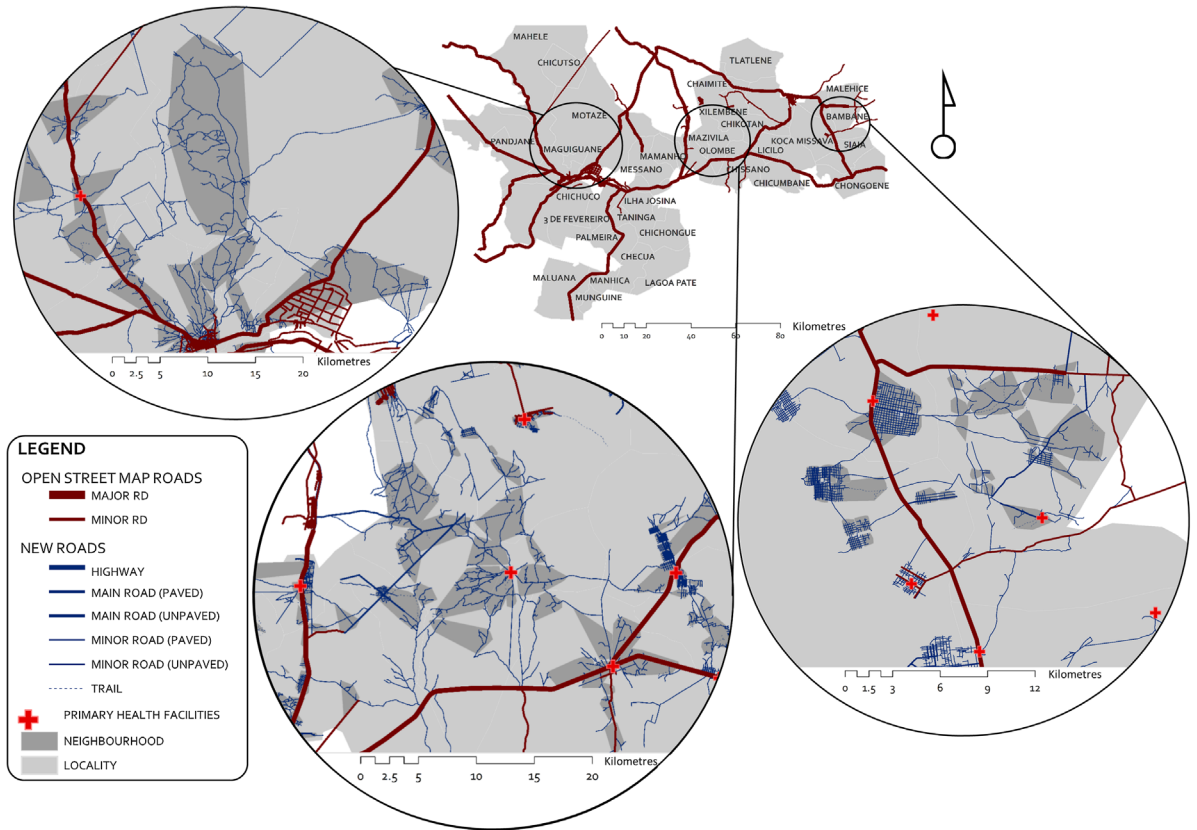
Figure 3
Data gaps in OpenStreetMap that were filled through manual digitizing.

up to the neighbourhood ID, so we would not be able to identify any of the individual households.

Based on this information we created Thiessen polygons around all GPS coordinates and dissolved the Thiessen polygons that had the same IDs for each of the three administrative levels (Figure 4). This process created two new sets of smaller community boundaries that did not exist prior to starting this work (in addition to the administrative post data): 36 localities and 425 neighbourhoods (Figure 5). While this approach created polygons around points that belonged to the same neighbourhood, locality, or administrative post, a key limitation of this approach is that unpopulated regions (e.g., forests) that lie in between the inhabited sections of neighbourhoods would be shared between these neighbourhoods. This approach,

therefore, tends to overestimate the spatial extent of the administrative units.

These neighbourhood boundary datasets were used for detailed spatial analysis of community-level access to care, in contrast to the more abstract possibilities that would have been achieved at the administrative post level. As each of the households were assumed to have one or more women of reproductive age (WRA), these data were also used to generate WRA population estimates which are an important component of evaluating access to maternal health services, and identifying marginalized populations of women.

A comparison of the official administrative post dataset with the ones created in this project revealed that 24 of the 425 neighbourhoods (or 2141 of the 50,619 households) in the study area
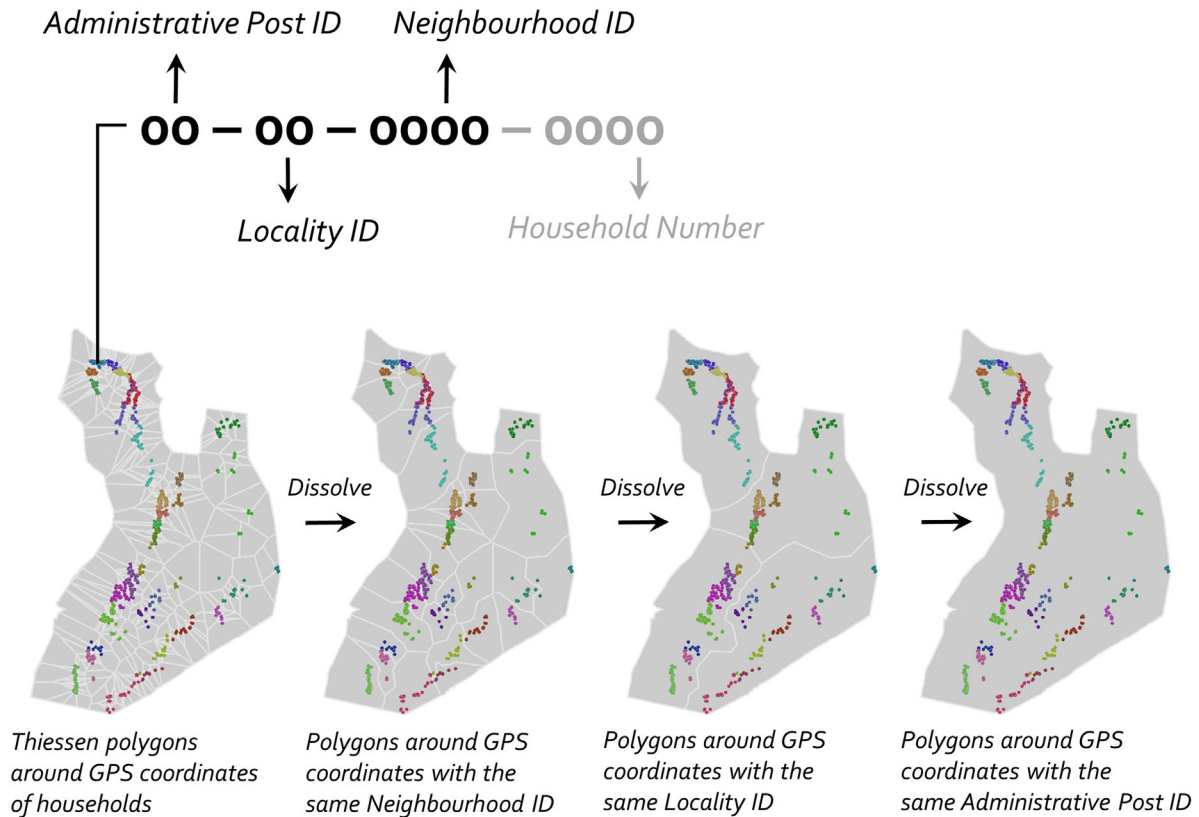
Figure 4
Structure of the household ID—used as basis for mapping community boundaries.

were not located within their presumed administrative post boundary by as much as almost 5 km in some of the cases.

Discussion

This article addresses some of the data challenges that health GIS researchers working in LMICs face and illustrates strategies to address these challenges by drawing on the lessons learned from our health GIS project in southern Mozambique. A set of guidelines that could help health GIS researchers in LMICs face these data challenges is summarized in Table 3. We found that although public databases are a good place to start searching for base GIS data, these sources do not provide enough data for high-resolution health-related spatial analysis. While

there is not an absolute lack of data in LMICs, decentralized and uncoordinated efforts result in duplication of mapping activities, as indicated by multiple custodians having different versions of the same dataset. It should be recognized that the institutions that are meant to coordinate creation and sharing of spatial data are not leading these initiatives in LMICs (Von Hagen 2007), which is a possible explanation for why most of the available data were acquired from public online sources rather than CENACARTA.

While much of the literature suggests that digitizing is a long, tedious, and expensive process (Sipe and Dale 2003; Awad 2013), the time taken to create the data from this work demonstrates that digitizing data is becoming cheaper and faster. This is attributed partly to the availability of free satellite imagery through the geoweb, which eliminates the
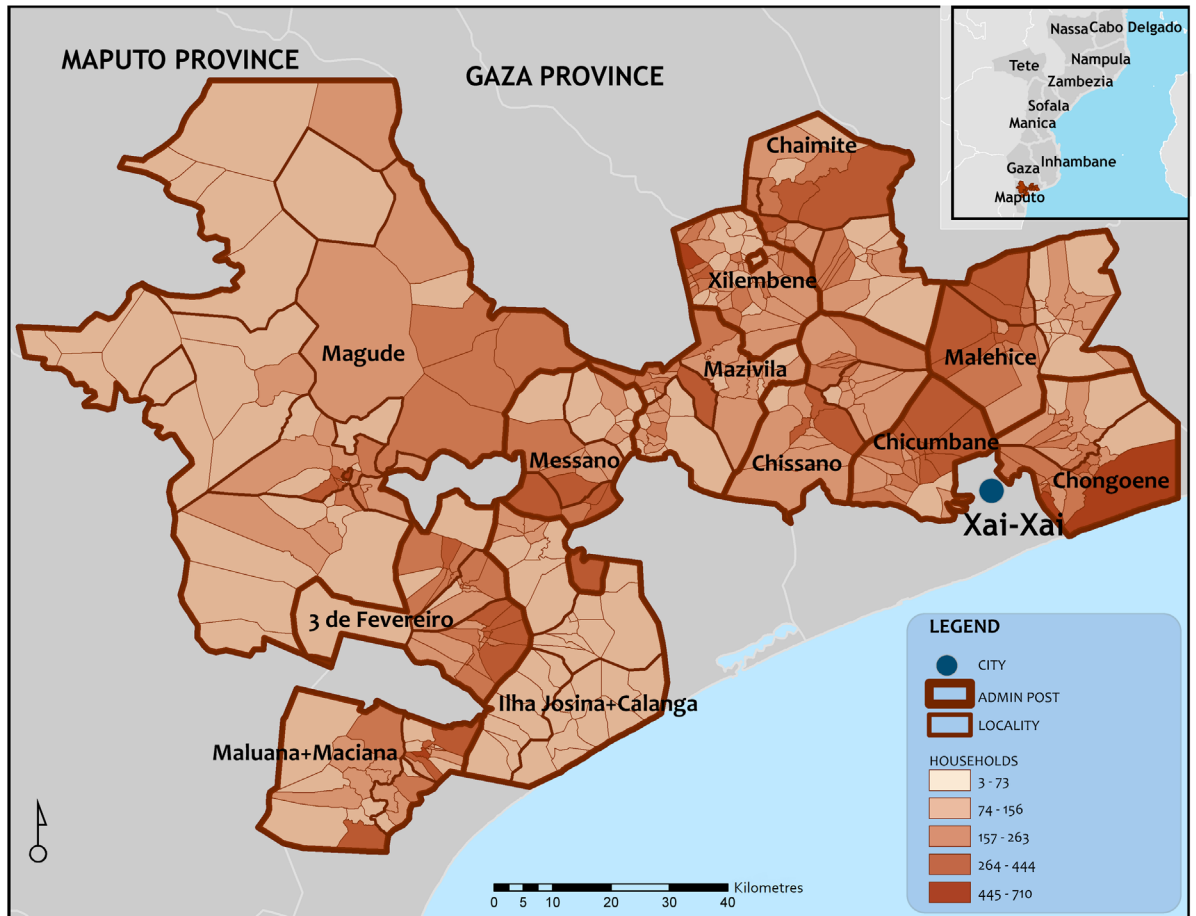
Figure 5
Community boundaries dataset.

high cost associated with their purchase. The distributed data capture, enabled through the use of open mapping standards like the WFS, makes the process more efficient. Manual digitizing may be perceived of as a trivial (and old-fashioned) method to many health GIS researchers, but it is an essential part of doing health GIS research at very granular levels in LMICs; it is also essential to sustaining the use of GIS in studying population health problems and targeting health interventions. Complementing the manual digitizing process with semi-automated methods may be helpful in some instances, although currently there are additional costs associated with these methods arising from post-processing and cleaning the data (Cao and Sun 2014).

The MOM data capture platform was designed using proprietary software because it was much easier and quicker to set up compared to other available open source options. The cost of acquiring this software may prohibit other LMICs from implementing a similar design, and we acknowledge this as a possible limitation of this work. However, web computing standards similar to the WFS are also available in most free and open source GIS software, and these could be used instead to achieve the same functionality as the platform presented in this article.

Table 3
Guidelines for gathering and creating framework GIS data in a typical data-poor setting

1. Determine what publicly accessible spatial data are available from existing data warehouses. This step precedes any decision to acquire base spatial data.
2. Consult local/national mapping agencies or other relevant mapping authorities and acquire datasets relevant to the research project. Inventory the data from all sources to identify data gaps.
3. Use freely available high-resolution satellite data to digitize new vectors that can be extracted from the imagery (e.g., dwellings and roads).
4. Utilize appropriately skilled local personnel with local knowledge to be part of a consensus-based process for data capture. Train all the data capturers how to interpret features from satellite imagery in a manner that is consistent.
5. Use GPS coordinates from previous household surveys or censuses, where available, as the basis for mapping the location of populations and higher-resolution community boundaries.
6. Utilize open geospatial standards for web mapping (e.g., the web feature service) to facilitate for distributed data capture, allowing for multiple users to work on the same centrally managed dataset.
7. Use open standards to document metadata for the created data; e.g., why the data were created, when, for what, by whom, use limitations, etc. This will allow for future users to know how and how not to use the data.
8. Use independent data checkers to validate the captured data for completeness, geometric precision, topological consistency, and classification of features. If possible, involve the mapping agency in this process.
9. Share the data with mapping agencies to add to their data infrastructure and to make data accessible to other researchers.

This project has produced high-resolution community boundaries that can be used for fine-grained spatial analysis that could help to better target health programs and inform health policy. It is likely that the generation of these community boundaries will be much faster and more accurate in the future because most countries in Africa are starting to use GIS and GPS as part of household surveys and censuses (Perez-Heydrich et al. 2013; Yilma 2015). Point data on household dwellings are also increasingly being used as an alternative to cadastre-based property registers (Hackman-Antwi et al. 2013; Statistics South Africa 2015), and this will open up opportunities to catalyze production of base community maps.

This project also demonstrates that the use of volunteered geographic information may not be ideal for the creation of framework data like roads. Data capturers for such high-precision geodata will require some level of training to achieve the required technical precision (Budhathoki et al. 2008). As was the case with the OpenStreetMap initiative, where only a small percentage of the registered users generate most of the content (Heipke 2010), our experience has shown that data will be captured more efficiently by a few trained individuals who have incentive to participate in the data capture process. Further to that, OpenStreetMap data for our study area, which had been created by other volunteers, was not precise enough for our work (Figure 3).

We undertook a participatory approach and involved CENACARTA in our data capture processes.

This enabled them to contribute to and validate the process of data creation. We anticipate that this instilled a sense of ownership in the data and its creation process. Intentionally involving members of CENACARTA also exposes them to basic retraining on methods of data capture that are much cheaper than those they currently use. As CENACARTA has the mandate to create and maintain framework spatial data (Rajabifard et al. 2006), working with this agency increases the potential for scaling up mapping work. However, more needs to be done to convert these methods into standard procedures that could be incorporated into routine mapping exercises by CENACARTA. Ultimately, it should be the role of the mapping agency to create and manage the fundamental GIS datasets and make them centrally accessible to different stakeholders, including health GIS researchers.

Conclusion

With the increasing inclusion of geographic thought and use of spatial techniques in health research, there is a growing demand for high-resolution framework spatial datasets. This article illustrates some of the key considerations for health GIS researchers working in settings where the required data may not be readily available. The processes presented in this article are a quick fix to the data challenges in most LMICs, and there is a need for more coordinated and sustainable efforts for data creation and sharing. Involving mapping agencies in

data capture processes has the potential to sustain rapid scale-up of mapping efforts, using some of the low-cost strategies presented in this article.

Acknowledgements

References

Aguilera, I., J. Sunyer, R. Fernández-Patier, G. Hoek, A. Aguirre-Alfaro, K. Meliefste, M. T. Bomboi-Mingarro, M. J. Nieuwenhuijsen, D. Herce-Garraleta, and B. Brunekreef. 2007. Estimation of outdoor NO_x, NO_2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. *Environmental Science and Technology* 42(3): 815–821.

Aimone, A. M., N. Perumal, and D. C. Cole. 2013. A systematic review of the application and utility of geographical information systems for exploring disease-disease relationships in paediatric global health research: The case of anaemia and malaria. *International Journal of Health Geographics* 12(1): 1–13. doi:10.1186/1476-072X-12-1.

Alford-Teaster, J., J. M. Lange, R. A. Hubbard, C. I. Lee, J. S. Haas, X. Shi, H. A. Carlos, L. Henderson, D. Hill, A. N. A. Tosteson, and T. Onega. 2016. Is the closest facility the one actually used? An assessment of travel time estimation based on mammography facilities. *International Journal of Health Geographics* 15(1): 1–10. doi: 10.1186/s12942-016-0039-7.

Awad, M. M. 2013. A morphological model for extracting road networks from high-resolution satellite images. *Journal of Engineering.* doi: 10.1155/2013/243021.

Bailey, P. E., E. B. Keyes, C. Parker, M. Abdullah, H. Kebede, and L. Freedman. 2011. Using a GIS to model interventions to strengthen the emergency referral system for maternal and newborn health in Ethiopia. *International Journal of Gynecology and Obstetrics* 115(3): 300–309.

Berendsen, M. E., J. D. Hamerlinck, S. G. Lanning, and S. Shin. 2010. *Developing NSDI framework data: ANSI framework data content standards guidance document.* Laramie, WY. Federal Geographic Data Committee: Wyoming Geographic Information Science Center, University of Wyoming.

Bing Maps. 2016. *Bing image analyzer.* http://mvexel.dev.openstreetmap.org/bing/.

Budhathoki, N., B. Bruce, and Z. Nedovic-Budic. 2008. Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal* 72(3-4): 149–160. doi: 10.1007/s10708-008-9189-x.

Cao, C., and Y. Sun. 2014. Automatic road centerline extraction from imagery using road GPS data. *Remote Sensing* 6(9): 9014–9033.

Ebener, S., M. Guerra-Arias, J. Campbell, A. J. Tatem, A. C. Moran, F. A. Johnson, H. Fogstad, K. Stenberg, S. Neal, P. Bailey, R. Porter, and Z. Matthews. 2015. The geography of maternal and newborn health: The state of the art. *International Journal of Health Geographics* 14(1): 1–10.

Ferguson, W. J., K. Kemp, and G. Kost. 2016. Using a geographic information system to enhance patient access to point-of-care diagnostics in a limited-resource setting. *International Journal of Health Geographics* 15(1): 1–12. doi: 10.1186/s12942-016-0037-9.

Fogliati, P., M. Straneo, C. Brogi, P. L. Fantozzi, R. M. Salim, H. M. Msengi, G. Azzimonti, and G. Pututo. 2015. How can childbirth care for the rural poor be improved? A contribution from spatial modelling in rural Tanzania. *PLoS ONE* 10(9): 1–18: doi: 10.1371/journal.pone.0139460.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2001. Scale issues and geographically weighted regression. In *Modelling scale in geographical information science*, ed. N. Tate and P. M. Atkinson. West Sussex, UK: Taylor and Francis, 123–140.

Friedman, J. M., L. Hagander, C. D. Hughes, K. A. Nash, A. F. Linden, J. Blossom, and J. G. Meara. 2013. Distance to hospital and utilization of surgical services in Haiti: Do children, delivering mothers, and patients with emergent surgical conditions experience greater geographical barriers to surgical care? *The International Journal of Health Planning and Management* 28(3): 248–256.

Gyamfi-Aidoo, J., C. Schwabe, and S. Govender. 2005. *Determination of the fundamental geospatial datasets for Africa through a user needs analysis: A synthesis report.* Tshwane, South Africa: Environmental Information Systems Africa.

Hackman-Antwi, R., R. Bennett, W. de Vries, C. Lemmen, and C. Meijer. 2013. The point cadastre requirement revisited. *Survey Review* 45(331): 239–247.

Heipke, C. 2010. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(6): 550–557.

Ivers, L. C., E. S. Garfein, J. Augustin, M. Raymonville, A. T. Yang, D. S. Sugarbaker, and P. E. Farmer. 2008. Increasing access to surgical services for the poor in rural Haiti: Surgery as a public good for public health. *World Journal of Surgery* 32(4): 537–542.

Kim, D., M. Sarker, and P. Vyas. 2016. Role of spatial tools in public health policymaking of Bangladesh: Opportunities and challenges. *Journal of Health, Population and Nutrition* 35(1): 1–5.

Makanga, P. T., N. Schuurman, P. von Dadelszen, and T. Firoz. 2016. A scoping review of geographic information systems in maternal health. *International Journal of Gynecology and Obstetrics* 134(1): 13–17.

Masters, S. H., R. Burstein, G. Amofah, P. Abaogye, S. Kumar, and M. Hanlon. 2013. Travel time to maternity care and its effect on utilization in rural Ghana: A multilevel analysis. *Social Science and Medicine* 93: 147–154.

MSDN (Microsoft Developer Network). 2016. *Understanding scale and resolution.* https://msdn.microsoft.com/en-us/library/aa940990.aspx.

Munyaneza, F., L. R. Hirschhorn, C. L. Amoroso, L. Nyirazinyoye, E. Birru, J. C. Mugunga, R. M. Murekatete, and J. Ntaganira. 2014. Leveraging community health worker system to map a mountainous rural district in low resource setting: A low-cost approach to expand use of geographic information systems for public health. *International Journal of Health Geographics* 13(1): 1–8. doi: 10.1186/1476-072X-13-49.

Nesbitt, R. C., S. Gabrysch, A. Laub, S. Soremekun, A. Manu, B. R. Kirkwood, S. Amenga-Etego, K. Wiru, B. Höfle, and C. Grundy. 2014. Methods to measure potential spatial access to delivery care in low- and middle-income countries: A case study in rural

Ghana. *International Journal of Health Geographics* 13(1): 1–13. doi: 10.1186/1476-072X-13-25.

Owoo, N. S., and M. P. Lambon-Quayefio. 2013. National health insurance, social influence and antenatal care use in Ghana. *Health Economics Review* 3(1): 1–12.

Perez-Heydrich, C., J. L. Warren, C. R. Burgert, and M. Emch. 2013. *Guidelines on the use of DHS GPS data*. DHS Spatial Analysis Reports No. 8. Calverton, MD. Measure DHS. https:// dhsprogram.com/pubs/pdf/SAR8/S AR8. pdf.

Rajabifard, A., A. Binns, I. Masser, and I. Williamson. 2006. The role of sub-national government and the private sector in future spatial data infrastructures. *International Journal of Geographical Information Science* 20(7): 727–741.

Richardson, D. B., N. D. Volkow, M.-P. Kwan, R. M. Kaplan, M. F. Goodchild, and R. T. Croyle. 2013. Spatial turn in health research. *Science* 339(6126): 1390–1392.

Schuurman, N., R. S. Fiedler, S. C. Grzybowski, and D. Grund. 2006. Defining rational hospital catchments for non-urban areas based on travel-time. *International Journal of Health Geographics* 5(1): 1–11. doi: 10.1186/1476-072X-5-43.

Shoff, C., T.-C. Yang, and S. A. Matthews. 2012. What has geography got to do with it? Using GWR to explore place-specific associations with prenatal care utilization. *Geo-Journal* 77(3): 331–341.

Sipe, N., and P. Dale. 2003. Challenges in using geographic information systems (GIS) to understand and control malaria in Indonesia. *Malaria Journal* 2(1): 1–8

Statistics South Africa. 2015. *Dwelling frame*. Page no longer active; last accessed May 30, 2016. http://mapserver2.statssa. gov.za/geographywebsite/Projects/dwellFrame.html.

Stevens, A. R., H. J. Onsrud, and M. Rao. 2005. Global Spatial Data Infrastructure (GSDI): Encouraging SDI Development Internationally. Paper presented at the 2005 Joint International Federation for Surveyors (FIG) and Global Spatial Data Infrastructures conference, April 16–21, in Cairo, Egypt.

Strobel, S, D. Sarafinof, D. Wesloh, and P. Lacey. 2016. *DGIWG — Web Feature Service 2.0 Profile*. Wayland, MA: Open Geospatial Consortium.

Tanser, F. C., and D. Le Sueur. 2002. The application of geographical information systems to important public health problems in Africa. *International Journal of Health Geographics* 1(1): 1–9.

UN ECA and DISD. 2001 (United Nations Economic Commission for Africa and Development Information Services Division). *The future orientation of geoinformation activities in Africa: An extended summary*. Addis Ababa, Ethiopia: UN ECA.

Von Hagen, C. 2007. Towards a spatial data infrastructure for Somalia using open source standards. *Journal of Spatial Science* 52(1): 157–169.

Yilma, A. D. 2015. Geospatially enabled census: Perspective for African countries. Paper presented at the Africa Geospatial Forum, August 18–19, in Johannesburg, South Africa.