Data-driven Local Planning at National Scale: How data collected on mobile phones enable a Conditional Grants Scheme in Nigeria

Prabhas Pokharel, Brett Gleitsmann, Salah Chafik, Candice Heberer, Myf Ma, Chris Tan, Zaiming Yao, Vijay Modi
The Earth Institute
Columbia University
New York, NY
Email: modi@columbia.edu

***Abstract*—Many countries struggle with effectively estimating what fraction of federal funds should be disbursed to various local administrative levels for spending on social programs for development. Varying spatial distributions of resources, populations, services, costs and past performance make local information vital. Traditionally, such data are collected using paper-based survey methods and aggregated up to larger admin units, introducing long delays and losing granularity of information.**

The authors assisted the Nigerian Office of the Senior Special Assistant to the President on Millennium Development Goals (OSSAP-MDGs) to rapidly collect data on health, education, and water facilities using Android smartphones. Data collection was facilitated by a generalizable platform (Formhub.org). Authors also built a common data visualization platform (Nigeria MDGs Information System) to present disaggregated data at facility and local government levels. These technological innovations were used by local planners to develop and evaluate more than $500M of federally-funded grant proposals that targeted more than 450 Local Government Areas over three years as part of the OSSAP-MDGs Conditional Grants Scheme. This paper presents the design and implementation of these systems in the context of enabling data-driven local planning.

I. Background

Spatial distribution of resources, populations, and services is not uniform; neither is the cost of service delivery. Cost effective allocation of resources and essential services requires planning level data at the local scale. Local actors understand local needs more than centralized funding agencies, but gathering local data poses a significant challenge. Traditional household (e.g., national census) and sector-specific facility inventories (e.g., Indian Minor Irrigation Census [1]) provide this level of detailed local data, but are time-consuming and expensive because of paper-based data entry and analysis. The slow turnaround time, typically measured in years, hinders the immediate planning needs of the country.

In recent years, mobile-phone connectivity and accessibility has greatly improved, making mobile data collection a real possibility across most of the world. Mobile data collection provides instantaneous access to survey results, improved management capabilities, and rapid turnaround time, among many other cost savings and efficiency measures. As a result, processes can be created by which data collected in the field can be easily integrated into common data-sharing platforms and accessed by actors at various levels of government. This allows the local-level planning processes to be more easily driven by recent and relevant data, than with traditional paper-based survey methods.

In 2005, the Nigerian Government secured debt relief from the Paris Club of Creditors, which effectively allowed Nigeria to annually reallocate roughly 150 billion Naira ($1B USD) previously dedicated to debt servicing to the commitment of achieving the Millennium Development Goals (MDGs). To oversee these newly available Debt Relief Gains (DRG) funds, the President created the Office of the Senior Special Assistant to the President on Millennium Development Goals (OSSAP-MDGs). Initially, funds were distributed directly to relevant federal Ministries, Departments and Agencies (MDAs) via negotiations between technical staff to enhance existing MDG-specific programs [2].

In 2007, OSSAP-MDGs created an innovative funding mechanism, the Conditional Grants Scheme (CGS), at the state level. The CGS was designed to stimulate partnership between the federal and state levels to involve state-level planners and to amplify the impact of federal funding transfers by leveraging additional funding from states. In this scheme, DRG funds are made available to states that provide 1:1 matching funds and OSSAP-MDGs retains final authority for approval of the MDG-specific grant proposals. The primary focus of CGS funding during the initial years has been for health, education and water interventions. As confidence in the scheme increased, federal spending on CGS increased from 18 billion Naira in 2007 (12% of the total OSSAP-MDGs budget of 150 billion Naira) to 32 billion Naira in 2011 (21%). The successful implementation of CGS at the state level generated the impetus to expand the scheme to the third tier of government, the Local Government Area (LGA) [2] [3].

In 2011, the CGS was expanded to include an LGA track, and funneled approximately 11.3 billion Naira to 113 LGAs in its first year. The Conditional Grants Scheme to Local Government Areas (CGS-LGA) requires the same stringent grant proposal process and 1:1 matching funds as that of the state track. Due to the much greater number of LGAs (774), the oversight necessary for the LGA track is much greater than it is for the state track. As a result, OSSAP-MDGs has hired additional management at the Federal level to oversee the

CGS-LGA, as well as 100+ LGA-based Technical Assistants (TAs) to support the program at the local level. The TAs play a critical role in improving connectivity among the three tiers of government, even beyond the CGS. They are also chiefly responsible for the development of the CGS-LGA proposals which form the foundation of the MDG-specific interventions [4].

The authors are a part of the team at the Earth Institute at Columbia University, which has partnered with OSSAP-MDGs to advise the design and implementation of the CGS-LGA program. The CGS-LGA program contains many innovative approaches in its implementation, including the selection of LGAs to receive funding on a pro-poor basis, incorporation of both data and local consultation in the proposal writing process, a multi-sectoral approach to assessing need, and intensive training provided to the TAs developing local level funding proposals. In this paper, we will focus on describing on how data played a role in the CGS-LGA planning process.

The primary motivations for making the CGS-LGA planning process data-driven are two-fold. First, many spatial factors affect the cost and the desired nature of service delivery; planning data at the local scale helps to cost-effectively allocate resources. However, much of the publicly available data in developing country contexts like Nigeria are aggregated to higher levels like Nigeria's states or the country itself. Second, while decentralized planning is more likely to reflect ground realities because planners are closer to communities facing the problems, simply leaving planning to local authorities creates substantial risk of capture of resource allocation process by local elites. One technique to abate this and increase accountability is to create a shared platform where local planners and central reviewers can access the same information about existing local resources.

To meet these needs, OSSAP-MDGs and the Earth Institute (EI) have created an online, geo-referenced database of health, education and water facilities in Nigeria. The platform is designed to serve the needs of two key groups: the TAs (local planners), and federal-level staff charged with reviewing and approving CGS-LGA proposals submitted by the TAs. Data for this system was collected via a baseline facilities inventory of health, education, and water points between 2011 and 2014. In the rest of the paper, Section II describes the design decisions behind the data collection and data usage systems, Section III describes the results this system has enabled; Section IV discusses key lessons learned, and Section V presents conclusions and recommendations for similar projects. The benefits of using data to inform the local planning process are not quantitatively evaluated in this paper.

II. System Design

As described, the authors set out to design a system that allowed for data from Nigeria's health, education, and water facilities to be used by two sets of users: local planners who would use the system to help design interventions to increase access to these services, and federal-level reviewers who provided feedback on these proposals and ultimately approved them. Given that the data didn't exist yet, the system for collecting the data had to be designed concurrently. Moreover, for added flexibility and speed of iteration, two different systems were designed: one for data collection and one for data presentation and usage.

In designing these systems, we had an underlying need for the system to be open source and extensible, as the systems would eventually be handed over to the Nigerian government.

A. Data Collection

The data collection system for the facilities inventory was based on two key design parameters:

1) Disaggregated geospatial data at facility level;
2) Quick turnaround between collection and use for planning.

One of the key factors influencing planning at the local level is the spatial distribution of existing resources. Varying spatial distributions of populations, services, costs, and past performance influence the needs of a given community drastically. The data driving local planning needs to account for these distributions and so capturing the geospatial location of each facility was essential.

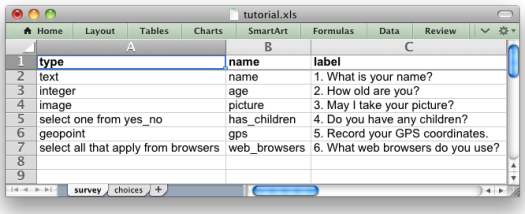
Moreover, the turnaround time on data collection needed to be fast. Due to the particularities of the national budget cycle in Nigeria, the first year of CGS-LGA plans were to be finalized and approved by February 2012. However, the first round of large-scale data collection (for 113 LGAs) was slated to take place in the summer of 2011 and so the data had to be collected, cleaned and available for use within a matter of several months.

Driven by these underlying needs, we chose to collect data using mobile phones. We used Android smartphones because of their low cost and high availability in the Nigerian market. They were also an ideal choice because most modern models contain embedded GPS devices, in addition to the ability to connect to the mobile phone network to submit data to our servers instantaneously. Given challenges of connectivity across much of rural Nigeria, survey data and GPS coordinates were collected while the phones were offline. Data was submitted once connectivity was available, sometimes at the end of the day and often at the end of a few weeks or a month.

The surveying itself was limited to only health, education, and water facilities. These sectors were chosen based on their relevancy to the MDGs and represent a manageable number of facilities for collection (compared to, for example, households). Moreover, understanding existing facilities provides a great way to judge the spatial distribution of existing service availability. In many ways, they provide us with a valid picture of the "supply side" of public services.

As for the specific technologies to design this data collection tool, we used the Open Data Kit suite's Collect application for collecting data on Android phones [5]. Collect's robust offline data collection capabilities, ability to capture GPS points, photographs, free-text responses and multiple choice questions, as well as the ability to include programmable flow control and constraint checks made it well-suited to our needs. In order to author forms for use in ODK Collect, as well as to aggregate the data, we introduced two key innovations.

The first innovation was XLSForm [6]. The ODK suite of applications is built upon the OpenRosa XForm specification,




```
<h:head>
  <h:title>tutorial</h:title>
  <model>
    <instance>
      <tutorial id="tutorial">
        <name/>
        <age/>
        <picture/>
        <has_children/>
        <gps/>
        <web_browsers/>
      </tutorial>
    </instance>
    <bind nodeset="/tutorial/name" type="string"/>
    <bind nodeset="/tutorial/age" type="int"/>
    <bind nodeset="/tutorial/has_children" type="select1"/>
    <bind nodeset="/tutorial/gps" type="geopoint"/>
    <bind constraint="(.='none' or not(selected(., 'none')))'
  </model>
</h:head>
<h:body>
  <input ref="/tutorial/name">
    <label>1. What is your name?</label>
  </input>
  <input ref="/tutorial/age">
    <label>2. How old are you?</label>
  </input>
  <upload mediatype="image/*" ref="/tutorial/picture">
    <label>3. May I take your picture?</label>
  </upload>
```

Fig. 1. Partial XLSForm (top) and XForm (bottom) representations of the same toy example survey. XLSForm is much more human readable, and has allowed for rapid iteration during survey testing and validation processes.

which is based on XML (Fig. 1). XML, or the Extensible Markup Language, is easy for machines to parse and understand, but significantly harder for humans to comprehend and author [7]. As a result, a simpler format for specifying surveys was needed. The spreadsheet format of Microsoft Excel documents was ideal for this process–each row was able to represent a single question, the question's type (free-text response, number, GPS point, etc.), as well as flow control and constraints pertinent to that question. The multi-sheet nature of Excel files also allowed for a compact representation of choices for multiple choice questions. Moreover, Microsoft Excel (and open-source variants) are pervasively used in the developing and developed world alike. XLSForm have made survey authoring for mobile data collection accessible to non-programmers, and increased collaboration and iteration for any team working with survey authoring using the OpenRosa specification.

The second innovation was Formhub.org [8], a server-side data collection platform. We developed Formhub.org to make the process of managing the data collection process easier. Formhub.org is built on top of XLSForm, and allows for a one-click map-based visualization of data being collected in the field. In turn, this visualization is available as soon as data is submitted to the Formhub server, which opens up new and transformative ways of monitoring the data collection process. Finally, Formhub couples the data collected to machine- and human-readable versions of the survey used to collect the data. This has allowed us to slowly automate away many of the data cleaning tasks that begin any data analysis and transformation process. Formhub also allows for data collected using the ODK collect platform to be downloaded by end users in multiple formats.

Data collection, however, is not the ultimate goal. The ultimate goal is data-driven planning, or the usage of data in a planning process.

B. *Data Usage for Planning*

For the CGS-LGA program, the data usage platform had the following key parameters:

1) Data highlighting aggregated facility-level data at the local government area (LGA) level;
2) Display of data for each facility that was surveyed in the inventory;
3) Planning-specific data for health, education, and water sectors (sectors funded by the CGS-LGA program);
4) Accessible to local planners as well as proposal reviewers at the Federal level.

The objective of the CGS-LGA program is to fund key interventions that help attain the Millennium Development Goals at the local level. In order to do this, local level planners should be able to use data to get a good glimpse at the status of MDG attainment in their local governments, as well as the status of health, education, and water facilities in the local government. As most of the interventions to be funded by the CGS-LGA were anticipated to be input-driven, we had to provide our local planners detailed information on the level of input accessibility within public service facilities in a given LGA.

While some of the interventions funded included the building of new health clinics or schools, more often than not, interventions focused around renovation or upgrade of existing facilities. To allow the TAs to plan well for this upgrade-based exercise, we needed to provide them detailed information at the facility level, as assessed by our survey. Facility by facility information is also crucial to reviewers at the federal level, who need to review the existing status of nearby facilities to judge the need for interventions to be applied at a given location.

Additionally, it is worth noting that presenting data as collected is not enough; we need to curate the data as it specifically helps inform the planning process for the health, education, and water sectors. For example, while all health facilities may be asked whether or not they perform cesarean sections for pregnant women, it is only hospitals and comprehensive health centers that must perform cesarean sections. As a result, the datapoint "Percentage of hospitals in local government area that perform c-sections" is much more relevant to the local planner than the "Percentage of all facilities in the local government that perform c-sections."

Moreover, the data has to be easily accessible to local planners and reviewers at the central level alike. At the local level, access to the Internet varies considerably, and some TAs are only able to access the Internet over satellite connections.

In order to meet these needs, we designed and constructed the Nigeria MDGs Information System, or NMIS (Fig. 2), [9]. NMIS is a website that provides rich access to the facilities

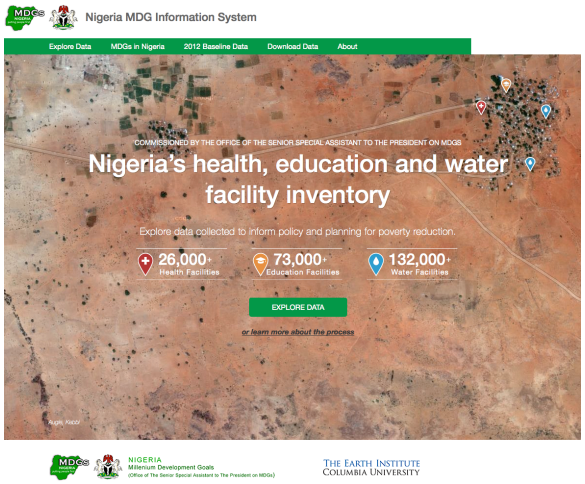


Fig. 2. NMIS homepage. Figures from NMIS are from a version that is not yet released to the public Internet.

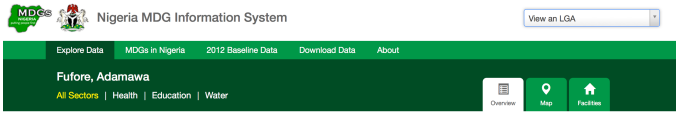


Fig. 3. Navigation in NMIS. Navigation within the data section is focused first on the LGA (Fufore LGA in Adamawa State has been selected in this example), then on the sectors (All Sectors, Health, Education, Water) and the type of data view (Overview, Map, Facilities). Only navigation is shown.

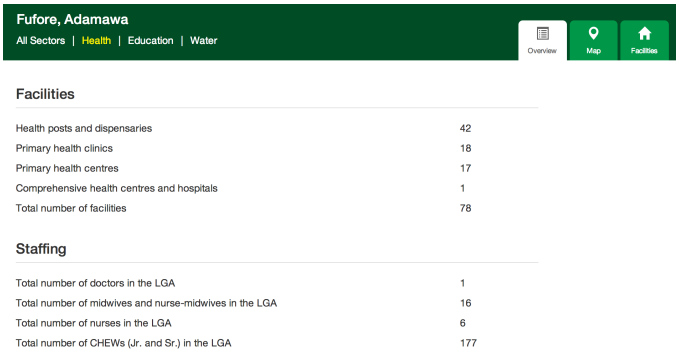


Fig. 4. Selected indicators from the Health Overview page. Main page navigation mitted.

inventory from a planning perspective, focusing on showcasing data at the LGA and the facility levels. To accommodate local level users with intermittent access to Internet, the website includes offline functionality; data about specific local government is downloaded onto a user's web browser and cached for future usage.

The information architecture of NMIS is focused around the LGA (Fig. 3). The main content of the website is present in the "Explore Data" section, the first action for which is to choose an LGA to view data in. On the LGA-page, there are two main navigational elements. The first navigational element lets users filter data by sector, including an "All sectors" option which presents a multi-sectoral view. The second element focuses on the way of viewing data, into Overview, Map,

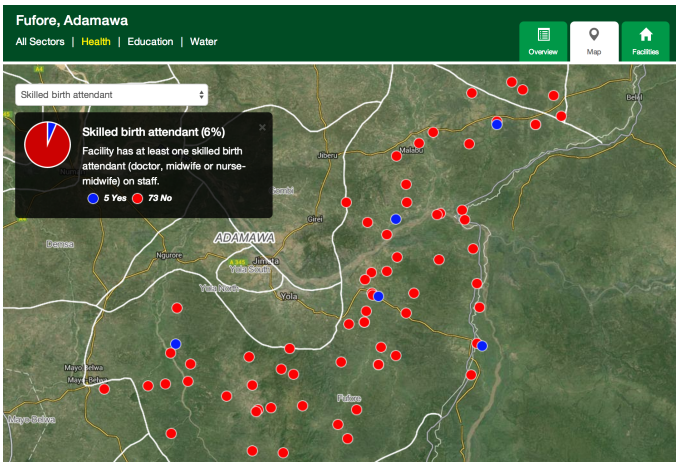
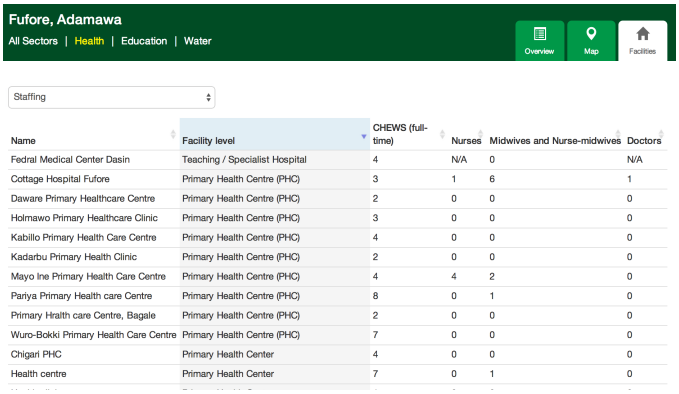


Fig. 5. A map-based visualization of whether or not health facilities have access to Skilled Birth Attendants. Main page navigation omitted.



Name	Facility level	CHEWS (full-time)	Nurses	Midwives and Nurse-midwives	Doctors
Fedral Medical Center Dasin	Teaching / Specialist Hospital	4	N/A	0	N/A
Cottage Hospital Fufore	Primary Health Centre (PHC)	3	1	6	1
Daware Primary Healthcare Centre	Primary Health Centre (PHC)	2	0	0	0
Holmawo Primary Healthcare Clinic	Primary Health Centre (PHC)	3	0	0	0
Kabillo Primary Health Care Centre	Primary Health Centre (PHC)	4	0	0	0
Kadarbu Primary Health Clinic	Primary Health Centre (PHC)	2	0	0	0
Mayo Ine Primary Health Care Centre	Primary Health Centre (PHC)	4	4	2	0
Pariya Primary Health care Centre	Primary Health Centre (PHC)	8	0	1	0
Primary Hralth care Centre, Bagale	Primary Health Centre (PHC)	2	0	0	0
Wuro-Bokki Primary Health Care Centre	Primary Health Centre (PHC)	7	0	0	0
Chigari PHC	Primary Health Center	4	0	0	0
Health centre	Primary Health Center	7	0	1	0

Fig. 6. Selected indicators from the Health Facilities page. Main page navigation omitted.

and Facilities. The Overview (Fig. 4) presents aggregated indicators at the LGA level (such as the total number of hospitals in an LGA, percentage of hospitals that perform c-sections in the LGA, etc.). The Map presents the various locations of health, education, and water facilities across the LGAs. In addition to clicking on a given facility to view its information, users can also view the spatial distribution of a given indicator (such as whether a facility has access to skilled birth attendants), so they can get a geo-spatial perspective of service distribution within an LGA (Fig. 5). Finally the Facilities view presents users a table where each row is a facility, and each column is an attribute of the facility, such as the facility's name, type, and for example the number of staff it has(Fig. 6). This final view allows local planners to get a detailed perspective of facilities within their LGA across any given dimension (such as the level of staffing across health facilities in that LGA).

At the LGA level, NMIS also contains "gap sheets," which present very specific data related to input availability in a given LGA. For these gap sheets, in addition to the level of service attainment, there is also a "target" or a minimum standard of input availability that the LGA should have, according to national standards, and a visualization of how far an LGA is from that target (Fig. 7). This tool offers a baseline framework

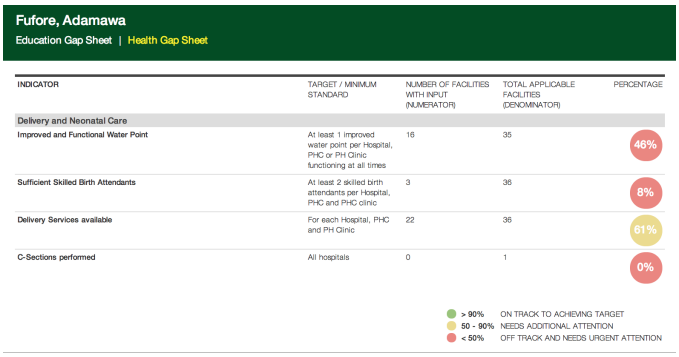


Fig. 7. Selected indicators from the Health Gap Sheet page. For these indicators, progress towards a target / minimum standard is visualized using traffic light colors.

for thinking about input availability in a given LGA, and has proven to be a popular tool in the local planning process.

III. Results

The system described above became a cornerstone of the CGS-LGA program and led to several key results:

1) Allocation of approximately $500M of federal funds to the LGAs via this conditional grant mechanism by mid-2014;
2) Creation of a nationwide database of over 250,000 health, education and water facilities (public access pending);
3) Formhub.org has spun off into a widely used, open-source data collection, management and visualization tool.

First, the systems described have enabled the CGS-LGA program to disburse approximately $500M of federal funds to be passed on to the local level, via three annual disbursements that have affected more than 450 LGAs. Throughout the development process, the data platform (NMIS) served as the principal planning tool for the CGS-LGA proposal writing process. TAs and LGA planners used the site data and data products (gap sheets) to prioritize interventions and develop grant proposals on an annual basis. These local planners were able to visualize their constituencies in new ways and make decisions based on geo-referenced facility-level data, which had never been previously done at this scale in Nigeria. In addition, the NMIS platform enabled OSSAP and EI technical experts and managers a way to concretely validate and evaluate grant proposals.

Second, a series of three major data collection efforts by OSSAP-MDGs led to the creation of the nationwide facility inventory of more than 250,000 health, education and water facilities. The dataset has been integrated into a user-friendly, web-based platform that TAs and LGA officials have been using since 2012. Local and international sector experts drove the survey content to provide detailed essential information for planning purposes, while the CGS focus on the LGA as the funding unit led to a data structure centered around the LGA. As of the time of this publication, the website is intended to be made public, which would make facility-level data available to a wide range of potential end-users from

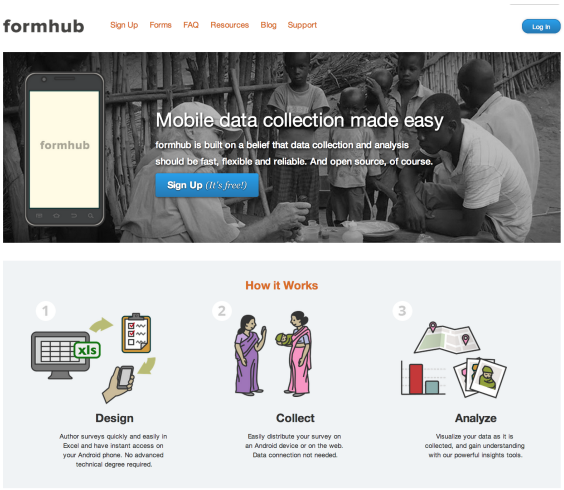


Fig. 8. Formhub.org homepage. The software is open source, and hosted freely for public usage. More than 7,000 users have collected more than 2.6 million individual surveys using the platform.

government to research institutions and civil society. Since data is downloadable from the website and includes broader geographic information (zone, state, etc.), NMIS users will have the ability to aggregate and organize data in ways that far exceed the relatively narrow goals of the data views on the website.

Third, the developers of the data collection tool (Formhub.org) intended from the outset to develop a product in the spirit of open-source. This tool was created to enable data collection with geocoding and offline capabilities specifically for the Nigerian context, but these criteria are broadly applicable around the world and as a result many other groups began to use Formhub. Currently, Formhub.org has more than 7,000 separate user accounts and more than 2.6 million survey forms have been collected, of which the Nigeria project accounts for less than 300,000. This unintended result highlights the importance of mobile data collection tools in these offline environments and is a development space that needs increased investment.

Beyond the direct outcomes, this data-driven process has been instrumental in improving the linkages between the LGA, State and Federal levels of government. The OSSAP-MDGs reports directly to the President but also in front of the National Assembly and in so doing, raises awareness on localized data-driven planning and encourages a paradigm shift. The TAs also play an important role in this effort. The simple fact that the LGA planning committee comes together to discuss NMIS data at least once a year during the grant proposal process can have far-reaching effects.

IV. Lessons Learned

The system described in this paper has now assisted local-level planning for three years. In that time, three distinct stages of data collection have taken place, along with three years of granting. Some lessons learned over this time are described below:

1) Enumerator capacity is one of the key determinants of data quality. Data quality absolutely depends on

the ability of the enumerator to understand the survey, convey that understanding to the respondent and to accurately input the data into the data collection device. When relying upon non-sector specialists as enumerators, these issues become even more paramount. Expenses linked to training and evaluation, testing and pilot exercises should not be underestimated and are absolutely necessary to ensure a minimum data quality standard.

2) Survey quality is as important for data quality; it is important to have focused but short surveys with simple and straightforward questions. When faced with the unique opportunity to conduct a "facility census", it can be tempting to collect as much data as possible when surveying facilities. Yet, shorter surveys can lead to less enumerator and respondent fatigue and therefore better data quality. The complexity of post-processing and data quality monitoring also grow with survey size. Simple and straightforward questions are necessary with non-specialist enumerators; guides on authoring effective surveys provide much assistance on this topic.
3) Ensuring good coverage of data capture is immensely difficult and important. To help local planners decide where a new hospital must go, existing data must describe all of the existing health services available in their community. One tool that is essential in this process are government-endorsed facility lists, which provide a measuring stick to know when surveying can be completed. Without such lists, one has to rely on local knowledge of where facilities are, which in turn leads to very inconsistent results and difficulty in evaluating how complete a dataset really is.
4) Data alone is not a panacea in the planning process. In the CGS-LGA process, data from NMIS is only one of the tools local planners use. They also confer extensively with the local community, the local government, and civil society organizations active in the community. TAs use data in addition to local knowledge, and are even allowed to present corrections to the NMIS dataset with documented evidence.
5) Mobile data collection effectively enables the local-planning process. The ability to capture GPS locations and photographs provides the context necessary when interpreting data; flow control and data entry constraint checks help improve data quality; the ability to submit data directly to the server reduces the time to produce usable, interpreted data; the ability to collect time-stamps and look at data immediately after submission creates new ways to monitor data collection and increase data quality. Moreover, with easy to use tools such as XLSForm and Formhub, non-programmers can become directly involved in the data quality monitoring and survey authoring process, which helps increase the ability to respond to lessons learned on the ground and improve the data collection processes.

V. Conclusion

In such a large national-scale, multi-sectoral project, there were obviously many challenges along the way. In the spirit of contributing to the global best-practice knowledge base, this section briefly summarizes several key recommendations and overall conclusions from the three-year effort.

The long-term success of the project will be largely determined by the degree to which the facility inventory dataset remains in use by local planning experts and how well it is kept up-to-date. In a word, this boils down to ownership. One point of potential weakness for the project in this regard is that it is housed within a special office to the president rather than within a permanent ministry, department or agency (MDA). This allowed us the flexibility to start quickly and actually use the data for planning, which may not have been the case had we gone through the more traditional MDA channel. However, the the fact that the OSSAP-MDGs office is not permanently anchored within government and that it can be viewed more as a political instrument, now leaves the dataset more vulnerable to neglect than if it were housed within a permanent MDA. As a result, the long-term future of the dataset, the local planning experts (TAs) and the updating mechanisms are uncertain and will be dependent on the continued existence and federal funding of the OSSAP-MDGs office. Updating requires a significant on-going financial commitment on behalf of the hosting agency and without this, the data will become quickly obsolete. Despite the use of mobile technology to collect and process data, facility-level visits are still required to obtain updated data. These visits require travel and logistical expenses to implement and therefore require long-term budgetary support to cover recurring costs for maintenance of the data.

And so, one key recommendation for similar large-scale data efforts is to structure the project in such a way that it is formally integrated into the permanent government structure. This will also help to ensure that MDAs endorse the final product. In this particular case, sectoral experts in education, health and water from the relevant MDAs were consulted along the way (e.g., survey authoring, enumerator training, data presentation), but nevertheless, the MDAs are thus far reluctant to fully endorse the dataset given that they were not ultimately responsible for the data collection, analysis and presentation. They are also unlikely to provide any funding for maintenance, updating or implementation as this remains firmly the responsibility of OSSAP-MDGs. Ideally, this effort would have been more formally embedded within the National Bureau of Statistics or other relevant MDA to properly institutionalize both the dataset and the data systems used for collection, analysis, and presentation.

The use of government-endorsed facility lists or similar datasets is also critical for credibility and official endorsement of the dataset. These lists provide data collectors a guide on what needs to be surveyed and also serve as an equally critical evaluation mechanism for assessing coverage. Collecting these lists can be difficult and time-consuming, but without this investment, the validity of the dataset will remain questionable. Engaging the local government at this level stimulates cooperation across the different levels of government, helps solidify a sense of ownership at the local level and generates buy-in of the final data product. Providing adequate resources

to obtain and vet facility lists is essential to the long-term success of the project.

The importance of focusing on enumerator capacity and survey clarity have been mentioned, but are worth underscoring as a final point as they underpin the data quality and therefore the credibility of the data set. Resource requirements for capacity building are dependent upon the expertise and background of the enumerator cadre and these need to be adequately evaluated from the outset. It is critical to understand the importance of this aspect to the project. The overall costs of the project are driven primarily by the costs of the actual data collection process (transportation and logistics, enumerator salaries, etc.). It is essential that the enumerators who undertake that process have the adequate training and survey instruments at their disposal to capitalize on the larger investment.

ACKNOWLEDGMENT

The authors would like to graciously thank the dedicated leadership and staff of the Nigerian Office of the Senior Special Assistant to the President on MDGs for their vision, collaboration, and funding that made this effort possible. We would also like to thank the Director of The Earth Institute and the entire Nigeria Scale-Up Initiative team for the support and confidence in our technical team to allow us to undertake such a rewarding project; the developers of the OpenRosa standard and the Open Data Kit suite for providing a powerful backbone for our platform; and the hundreds of local enumerators and TAs who have done all of the hard work on the ground to create this wonderful facility inventory.

REFERENCES

[1] "Indian Minor Irrigation Scheme", July 2014. [Online] Available: http://micensus.gov.in/CensusMISch.html

[2] C. Zamba and V. Oboh. "Effective Partnerships for Accelerating the Millennium Development Goals (MDGs) at the Sub national Level: Evidence from the Implementation of Nigeria's Conditional Grant Scheme (CGS)", 2013 Global MDG Conference, UNDP Publishing, December 2013.

[3] Office of the Senior Special Assistant to the President on Millennium Development Goals. "Partnering to achieve the MDGs: The Story of Nigeria's Conditional Grants Scheme 2007 - 2010", 2010.

[4] Office of the Senior Special Assistant to the President on Millennium Development Goals. "MDGs Conditional Grants Scheme Implementation Manual (Revised)." 2012.

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

[6] "XLSForm.org", June 2014. [Online]. Available: http://xlsform.org

[7] T. Parr. "Humans should not have to grok XML", IBM Developer Works, August 2001. [Online] Avaialble: http://www.ibm.com/developerworks/xml/library/x-sbxml/index.html

[8] "Formhub", June 2014. [Online]. Available: http://formhub.org

[9] "Nigeria MDG Information System", June 2014. [Online] Available: http://nmis.mdgs.gov.ng/