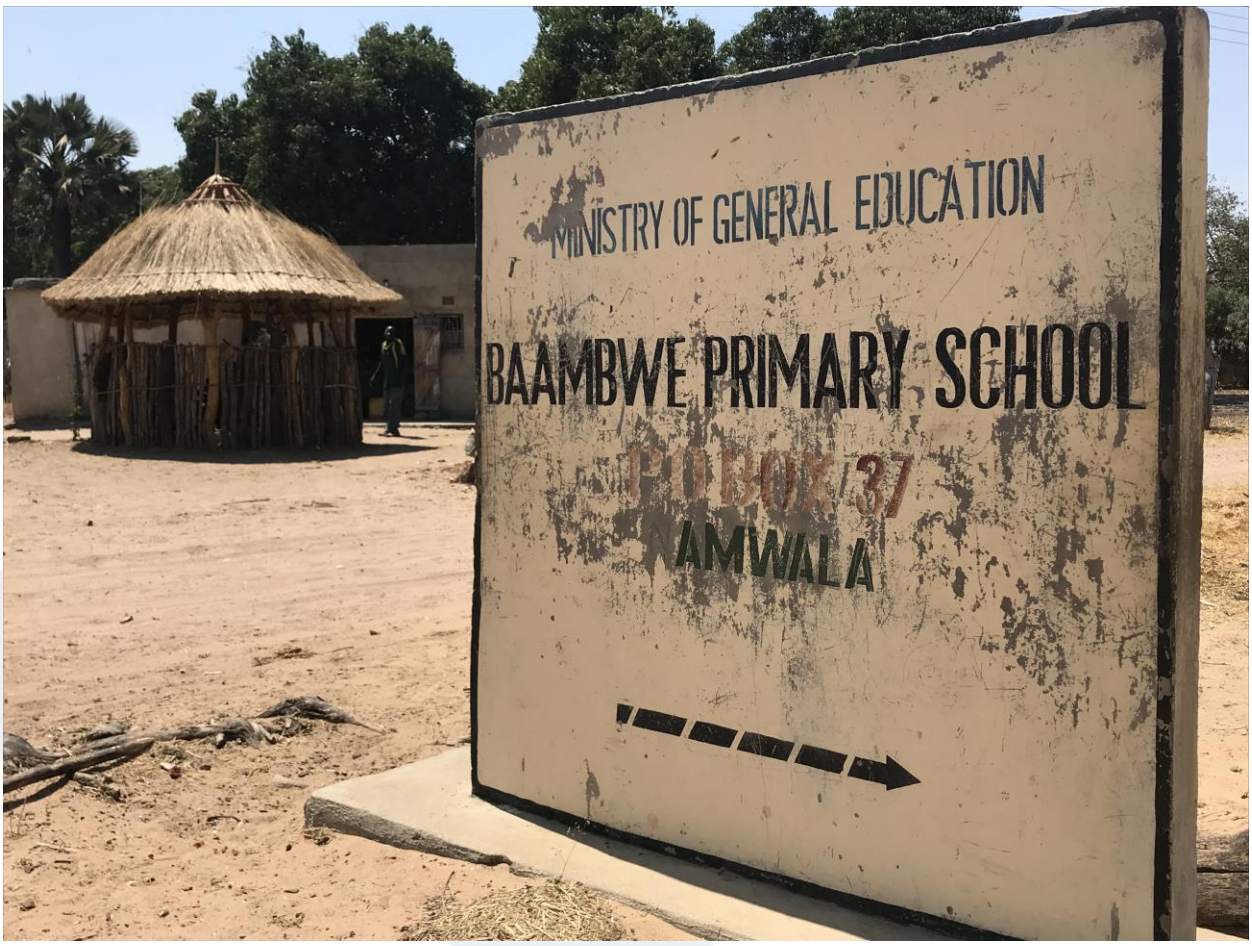


NANO DIY

A GUIDE TO DEVELOPING AN INFORMATION SYSTEM FOR TRADITIONAL LEADERS

05 December 2019

IDinsight
DATA. DECISIONS. DEVELOPMENT.

Authors

Valentina Brailovskaya: Valentina.Brailovskaya@IDinsight.org
Ana Chaves: Ana.Chaves@IDinsight.org
Christopher Chibwana: Chris.Chibwana@IDinsight.org
Eric Dodge: Eric.Dodge@IDinsight.org
Martin Gould: Martin.Gould@IDinsight.org
Kondwani Mumba: Kondwani.Mumba@IDinsight.org
Nate Vernon: Nate.Vernon@IDinsight.org
Paul Wang: Paul.Wang@IDinsight.org

About IDinsight

IDinsight uses data and evidence to help leaders combat poverty worldwide. Our collaborations deploy a large analytical toolkit to help clients design better policies, rigorously test what works, and use evidence to implement effectively at scale. We place special emphasis on using the right tool for the right question, and tailor our rigorous methods to the real-world constraints of decision-makers.

IDinsight works with governments, foundations, NGOs, multilaterals and businesses across Africa and Asia. We work in all major sectors including health, education, agriculture, governance, digital ID, financial access, and sanitation.

We have offices in Bengaluru, Dakar, Johannesburg, Lusaka, Manila, Nairobi, New Delhi, San Francisco, and Washington, DC. Visit www.IDinsight.org and follow on Twitter @IDinsight to learn more.

TABLE OF CONTENTS

1. INTRODUCTION

This do-it-yourself (DIY) guide describes how to build a user-friendly and decision-relevant information system (IS). The guide is based on an IDinsight pilot of the Nano IS in Zambia.

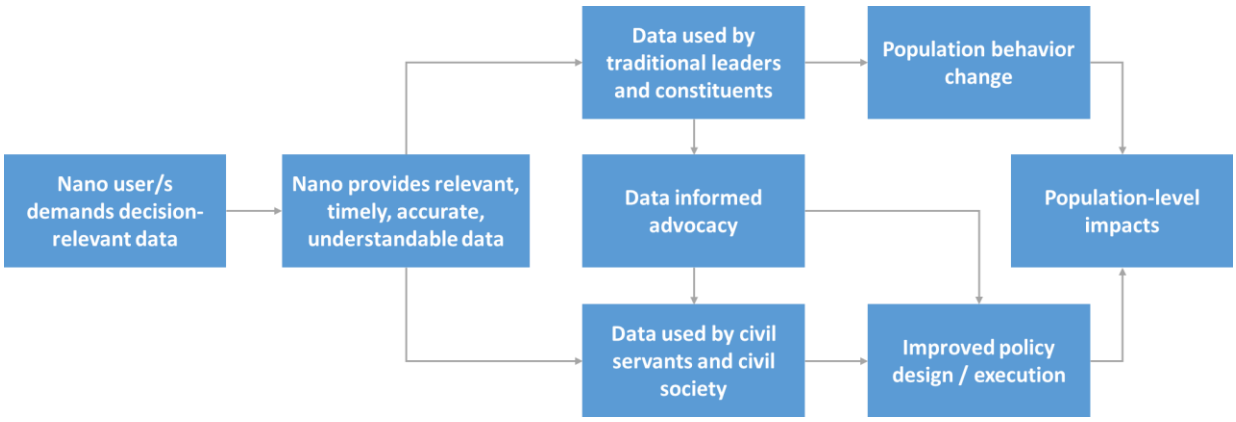
The purpose of the guide is to help organizations learn from the Nano project and to reproduce parts or all of the IS we built. It is best suited for individuals and originations with intermediate to advance research and data collection experience.

OVERVIEW OF NANO MOTIVATION AND PRINCIPLES

Efforts to implement information systems to enhance state capacity are widespread. However, these efforts often fail due to top-down, complex, slow, and expensive design and execution.[1] IDinsight is developing Nano as a solution to these issues – an aggressively consultative, low-cost, and flexible information system that comprehensively adheres to predictors of IS success.

Nano's theory of change (Figure 1) is that access to relevant, accurate, timely, and well-presented data that users demand will yield data-based actions to influence government, civil society, and population behaviour to improve lives.

Figure 1: Nano's theory of change



We expect that users will love Nano because it quickly produces critical and decision-relevant information they demand in an intuitive and engaging format. Nano has five components (see Table 1), and we have developed each of these in a pilot in Zambia.

[1] Hastie, S., & Wojewoda, S. (2017, July 19). Standish Group 2015 Chaos Report - Q&A with Jennifer Lynch. Retrieved from InfoQ: https://www.infoq.com/articles/standish-chaos-2015/

Table 1: Nano components

Component	Description
1. Consultation and stakeholder buy-in	Process by which we identify needs/demand and gain the support of all relevant stakeholders
2. Survey modules and data collection modes	Library of survey modules for streamlined data requests, tailored to the manner in which data can be collected (household interviews, service provider interviews, observations)
3. Sampling	Sampling strategy and sample size determination process
4. Enumerator management	Recruitment, training, and management (in-field and remote) including incentives/remuneration
5. Data presentation	User-friendly interface with data reports and guidance on how to act on data

DESIGN OF THE PILOT IN ZAMBIA

Nano was piloted in Mukobela chiefdom, Southern Province, Zambia[2], supported by the Global Innovation Fund.

Traditional leaders are important decision-makers, and policymakers in Zambia.[3] As locally embedded and widely respected leaders, chiefs are uniquely able to mobilise collective action, bypassing common challenges faced by government counterparts. However, they lack access to basic information on their chiefdoms and so, are often unable to identify and fully address issues for their communities.

Information systems implemented for decision-makers often struggle to fill this gap due to their complexity, cost, unwieldy design, and lack of consultation and context specificity.[4] Nano attempts to solve these problems through demand-driven design, systematically identifying and incorporating traditional leaders' information needs at each stage of the design process. It does this through frequent consultation with the end-user and design iteration.

A paucity of decision-relevant information is not exclusive to Southern Province or traditional leaders, affecting actors across domains in developing countries.[5] The learnings from the Nano pilot in Zambia are applicable to other levels of government across Zambia and governance structures in other developing countries.

[2] Mukobela Chiefdom is located in Namwala District, Southern Province. The district has a population of 102,866 across four chiefdoms. Mukobela Chiefdom is approximately 807 km2, part of which includes the district centre and central business district. Mukobela Chiefdom has eight village clusters and approximately 130 villages.
[3] Koenane, M.L.J. (2018). The role and significance of traditional leadership in the governance of modern democratic South Africa. Africa Review, 10:1, 58-71.
[4] Hastie, S., & Wojewoda, S. (2017, July 19). Standish Group 2015 Chaos Report - Q&A with Jennifer Lynch. Retrieved from InfoQ: https://www.infoq.com/articles/standish-chaos-2015/
[5] OpenData Barometer. (2016). ODB Global Report Third Report. The World Wide Web Foundation.

PURPOSE OF DIY GUIDE

This DIY guide describes the process of building the Nano information system. The purpose of the guide is to help organizations with research and data collection experience to reproduce the Nano project.

The guide is structured as follows:

1. Stakeholder and user consultation
2. Module selection and design
3. Sampling
4. Enumerator management
5. Data presentation

2. STAKEHOLDER AND USER CONSULTATION

OVERVIEW OF THE CONSULTATION PROCESS

Developing an IS involves two levels of consultation:

1. stakeholder consultation to facilitate ease of work when approval from relevant departments is required, and
2. user consultation to ensure user needs/preferences are reflected on information collected, process decisions, and interface features.

The consultation process begins with stakeholder engagement to get relevant approval and identify potential users and key contact people. Consultation with a wide range of stakeholders allows for the identification of areas of cooperation and overlapping information needs between the primary stakeholder (Chief) and other partners.

Stakeholder engagement is followed by user consultation, which takes place repeatedly throughout the IS development process to ensure that the product appropriately incorporates user needs and preferences. Consultation with users and stakeholders creates an opportunity to leverage already existing data and local knowledge and allows for continuous adjustment and improvement of the IS.

This section describes our approach to consultation and provides guidelines for engaging with government departments and traditional leaders.

STAKEHOLDER CONSULTATION

Working with government partners

Government approval is typically required before beginning any engagement. Buy-in from the government is necessary in order to demonstrate respect for existing actors in the space and facilitate ease of work later in settings that require government permissions.

Government departments adhere to set processes and procedures with regard to communication and stakeholder engagement. Budget additional time to account for government processes as these may result in delays. In the case of Nano, it took between four and eight weeks to get written approval from government partners.

Government engagements have to be pre-approved by a central authority, who is usually the Permanent Secretary in the relevant ministry for your sector/ stakeholder. Seek this approval in writing from the ministry and request an in-person meeting to provide context and clarification about the project. When organizing such meetings, involve other ministry staff like department heads, as they are usually more familiar with the ministry's operations and are more likely to better understand how your project will benefit the ministry. They may be a valuable ally in getting buy-in of the approving authority (Permanent Secretary).

For Nano, we sought approval from the Ministry of Chiefs and Traditional Affairs (MOCTA), as our primary stakeholder was Chief Mukobela. We wrote to the Permanent Secretary, expressing our interest in engaging with them and requesting an in-person meeting. After two in-person meetings, MOCTA granted approval for the project and directed us to the appropriate provincial and district

contact people. We obtained written confirmation of the ministry's support and submitted this to the ethics board assigned to our project to demonstrate engagement and buy-in from the appropriate government counterpart.

Starting work in a new district

When starting work in a new district, familiarize yourself with the hierarchy of district administration and pay courtesy calls to relevant district administration officials. This usually takes the form of brief (~30 minutes) visits to each official responsible for the facilities you intend to visit or the information you intend to collect.

Contact the district official assigned to you at least a week prior to your visit to organize meetings with local stakeholders. Create an itinerary with proposed times for meetings with different stakeholders and allow them time to schedule these meetings.

For Nano, we worked with the District Chiefs and Traditional Affairs Officer (DCTAO). He scheduled meetings with the District Commissioner, District Council Secretary, and heads of departments from the agriculture, education and health ministries.

USER CONSULTATION

Working with a chief

When working with a chief, always go through the government department responsible for traditional leaders. These officials will brief you on norms and expected dress and conduct. On your first meeting with the chief, you may be required to present a gift to the chief, and you should ask the government official for guidance on norms for this gift.

The government official should arrange the first meeting with the traditional leader, brief the chief on the nature of your visit, and accompany you to the meeting to demonstrate the legitimacy of your visit. For Nano, the DCTAO arranged a meeting with Chief Mukobela. We were accompanied by staff from a partner NGO, the Ministry of Health, and MOCTA.

What to consult on

User consultation covers a wide range of topics from survey content to process improvements. The following list, though not exhaustive, describes useful topics to discuss with the chief:

1. Chiefdom administration: Familiarize yourself with administration structures and information flow in the chiefdom. This will allow you to identify information gaps that your IS can fill and the appropriate administrative units at which to collect and present data.
2. Survey content: Ask the chief to identify modules that would be of interest to him. Then, select modules from this list by considering how actionable and feasible each module is (Section 1 provides details on evaluating and selecting modules).
3. Indicator selection: After module selection, ask the chief to identify specific indicators in each module that would be most actionable. These indicators should include specific statistics that the chief is interested in, e.g. pupil-to-toilet ratio.

4. Interface design: Design a prototype interface using example data and present it to the chief. Solicit feedback on how easy the interface is to use and understand. Iterate this process several times to develop an informative and intuitive interface.
5. Pathways to action: After presenting data to the chief, ask about potential actions for each indicator and stakeholders required to take action. Doing so provides insight into the relevance of the data collected and learnings for the design of future modules.

BEST PRACTICES WHEN MEETING STAKEHOLDERS

To successfully design an IS that users love, remember the following:

1. Formalize consultation: Consultation may take the form of formal meetings or semiformal conversations with users and stakeholders. Design brief interview guides in advance of stakeholder meetings to elicit information in a systematic and comprehensive manner. At minimum, each interview guide should cover the following: permissions and information needed to engage with the stakeholder's domain, current priorities and challenges, and existing programs or processes that should be considered when planning the IS process. An example interview guide is included in the Appendix.
2. Consult with a wide range of stakeholders: Engaging with a broad set of stakeholders enables you to leverage already-existing data and local knowledge, and helps to prevent duplication of efforts. Ask stakeholders to introduce/ lead you to other parties who could be relevant to your work as they are usually aware of other partners working in their district/ region.

3. MODULE SELECTION AND DESIGN

OVERVIEW OF MODULE DESIGN AND SELECTION

Meeting user's data needs requires designing survey modules that collect data that is decision-relevant and easy to understand. This section describes the process of module selection and design and provides insights on appropriate data collection tools.

MODULE SELECTION

Module selection should be driven primarily by the chief's preferences, and secondarily, by other factors that influence social impact like interest from other stakeholders, ease of administration, and pathways to action. Pathways to action may require support from other stakeholders; consult these stakeholders to determine what is most useful and avoid duplication of efforts. Design a structured matrix to assess each module on these factors. This matrix should include measures on decision relevance, interest from the chief and other stakeholders, feasibility, ease of training, and path to action. An example matrix is given in the Appendix.

When discussing potential modules with the user, set realistic expectations on the type of data you can collect and the administrative units at which results can be presented (i.e. presenting village, zone, village cluster, or chiefdom level statistics). Some requests, such as a chiefdom census or village-level statistics, may not be feasible due to time, cost, and privacy constraints. Reporting village-level data, for example, may be more costly and more time-consuming and jeopardize the anonymity of respondents in small villages.

MODULE DESIGN

The goal is to design modules that are usable across a wide variety of regional, country and user contexts. This can be achieved by basing questions on standardized HH surveys (e.g. the Demographic and health survey, Rural Agriculture and Livelihoods Survey, the national census, etc.). Results should translate to intuitive and informative visualizations, so options for future data presentation should be considered when designing survey questions.

Translating survey questions to local languages can potentially change the meaning and introduce noise in the data. Questions with complex sentence structures are harder to translate and should be avoided. Use local translators to ensure that translations adhere to local dialect and slang.

Data cleaning and analysis can be time-consuming, and modules should be designed to reduce the amount of effort required for these tasks. Keep open-ended questions to a minimum as they require more effort to clean.

When designing survey modules, remember to include the following:

1. Respondent identification and contact information. This should include names (both official and common nicknames) and contact information to allow you to locate respondents in the future.
2. Household GPS coordinates, descriptions, and directions to locate households on subsequent visits.

3. Time audits to make sure enumerators are administering the survey correctly. These can be added to measure the duration of the entire interview and individual sections or questions.
4. Text audits to check if respondents are changing responses to influence survey length. Text audits keep a record of original response recorded and are useful for questions that trigger sections of the survey and influence interview length.
5. Audio checks to ensure enumerators are asking questions and recording responses correctly. Adding text and audio audits requires additional files to be uploaded to the server with each survey. This might be challenging in places with a poor internet connection. Keep these to a minimum to avoid disruptions in survey uploads.

USING THE RIGHT TOOLS

Plan to collect data using smartphones or tablets, in order to allow for logic checks, skip patterns, and validations during surveying. This also allows for survey monitoring through automatically recorded timestamps, GPS locations, and text and audio audits. Depending on the internet connection, surveys can be accessed immediately after the interview is completed, facilitating high-frequency checks.

There are several software options available to enable data collection via tablets. When deciding which to use, consider the following:

1. **Price**: The cost of software ranges from free to one-time fees and monthly subscription. The project budget can be a limitation to which software you use.
2. **Required hardware**: Software runs on mobile phones, tablets, or computers. Software that runs on computers requires larger capital investments in equipment and may be limited by battery life and connectivity.
3. **Programming experience**: Some software requires advanced programming skills. Select software that is suited to the programming experience of your in-house staff. While outsourcing programming is a viable option, it is less desirable as it may increase project costs and reduces the capacity to adjust survey content as need arises.
4. **Data quality monitoring options**: Evaluate the feasibility of quality checks based on region-specific constraints and use this as a guide to selecting appropriate software. A poor internet connection can make it difficult to upload audio files, and cloud cover (common when surveying during the rainy season) can reduce the GPS capabilities of mobile devices.

4. SAMPLING

Sampling involves selection of a subset of the population in the study area for surveying. Sampling can provide relatively accurate estimates of a populations characterises without the need to survey the entire population. This section describes sampling approach used in the Nano project.

OVERVIEW OF SAMPLING

Project areas vary in size and can have over 1,000 households spread across a hard-to-access geography. Surveying all households in a study area is not feasible when building a fast, cheap IS. This necessitates the development of an up-to-date list of households (or villages) that could be used as a sampling frame (from which to draw a sample). A good sampling frame should be updated, accurate, and representative at the decision-relevant level (e.g. village, chiefdom, ward, or district). In this section, we explore three sampling frame options, i.e. census, administrative data, and geographic sampling, and give step-by-step details on geographic sampling.

Census

This entails carrying out a one-off census of the study area and using this as a sampling frame for subsequent surveys. A census is a complete record of the study population with basic demographic information, so it allows for stratification of specific subpopulations like girls and school-going children.

While ideal due to completeness and possibility of stratification, conducting a census is costly and time-consuming, and this is increased by the logistics of reaching every household in a rural chiefdom or population. Further, population changes due to migration into and out of the study area, birth, and death necessitate periodical updating of census records.

Administrative data

Two sources of administrative data are census estimates and village registers. These sources save on cost and time as they are usually accessible through the census bureau or traditional leaders, such as the chief and village heads.

Census data is periodic (every 10 years) and, therefore, usually outdated. Additionally, it is not always disaggregated at the right level. For Nano, census data was disaggregated up to the ward level, while we needed data disaggregated at the village cluster level.

Village registries maintained by village heads are usually outdated, may contain incorrect entries, and are sometimes non-existent. For Nano, we accessed master village registers kept by the chief, some of which had not been updated in the last 5 years and had households incorrectly entered,[6] resulting in an overestimation of the population.

[6] Some individuals belonging to one household were incorrectly listed as separate single-person households, inflating population estimates.

Geographic sampling

Geographic sampling involves building a sampling frame using population data and digital maps. For the Nano project, we achieved this by partitioning the chiefdom into small non-overlapping geographical units (enumeration areas).

Geographic sampling can be done ex-situ with open source software and publicly available population data. It presents the advantage of being fast, cheap, and reliable. In this section, we describe how we used geographic sampling in the Nano project.

COMPARING THE THREE SAMPLING METHODS

A sampling method suitable for an information system like Nano should be cheap, fast, and accurate. A census creates an accurate sampling frame, but at the trade-off of cost and time. While administrative data can be used to create a sampling frame quickly and at low cost, the levels of disaggregation, completeness, and age of the data make it undesirable. Geographic sampling meets all three criteria, as it can be quickly done with freely available data that is regularly updated by entities like Facebook and Openstreetmaps. Some strengths and drawbacks of each method are summarized in the table below.

Method	Cost	Time taken	Accuracy	Source of bias
Census	High	High	Very accurate	Measurement and non-response
Administrative data	Low	Low	Moderately accurate	Disaggregation level, age of data
Geographic sampling	Medium	Medium	Accurate	Exclusion error from different rooftop/ building identification techniques

GEOGRAPHIC SAMPLING PROCESS

Step 1: Plotting all the shapefiles on the map

Start by plotting the boundary for the study population. In our case, we had access to a chiefdom's boundary shapefile. Next, include any other shapefiles that may serve as natural borders to smaller segments (roads, waterways, mountains — whatever is relevant and available).

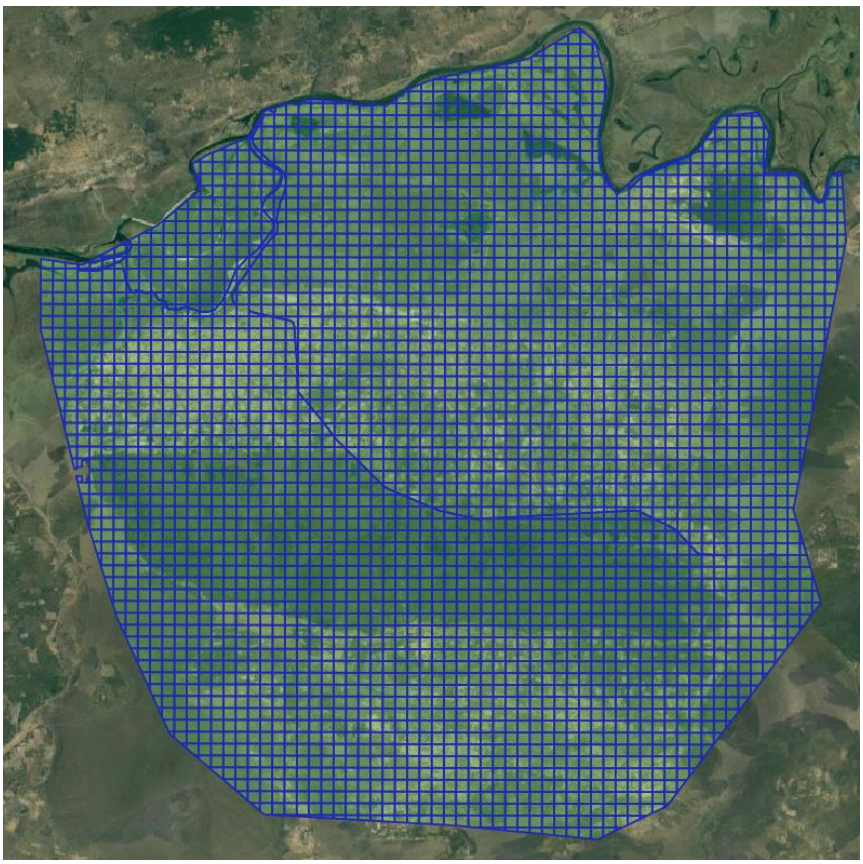
Humdata[7] is an good source for all sorts of datasets (some of the datasets are not complete but having some information is useful).

Step 2: Segmentation of the full study area into enumeration areas (EAs)

Segment the chiefdom into smaller non-overlapping areas (500x500 meter grids in our case). The size of the EA was somewhat arbitrary, but the goal is to ensure that they are small enough to be manageable for the surveyors on the ground, but large enough to contain two households on average. Then, trim each EA to natural borders to make it easier for surveyors to stay within EA boundaries by avoiding natural landscapes or infrastructure splitting an EA.

[7] https://data.humdata.org/

Figure 2: Segmentation of Chiefdom into EAs



Step 3: Determining areas with a high probability of household presence

Identify EAs with some indication of household presence as portions of your study area might be uninhabited. This is where Facebook's population and OpenStreetMaps rooftop datasets are useful (hereafter called "FB" and "roofs" dataset, respectively).

The advantage of using both datasets was that they were constructed differently. Facebook used a machine-learning algorithm to identify populated areas, whereas the OpenStreetMaps dataset was produced by a volunteer-based structure-tagging exercise. If each dataset has systematic exclusion errors, the reasons for exclusion are likely different between the two. Therefore, using both datasets was more likely to result in a sampling frame without systematic exclusion of certain types of households, which would have introduced bias.

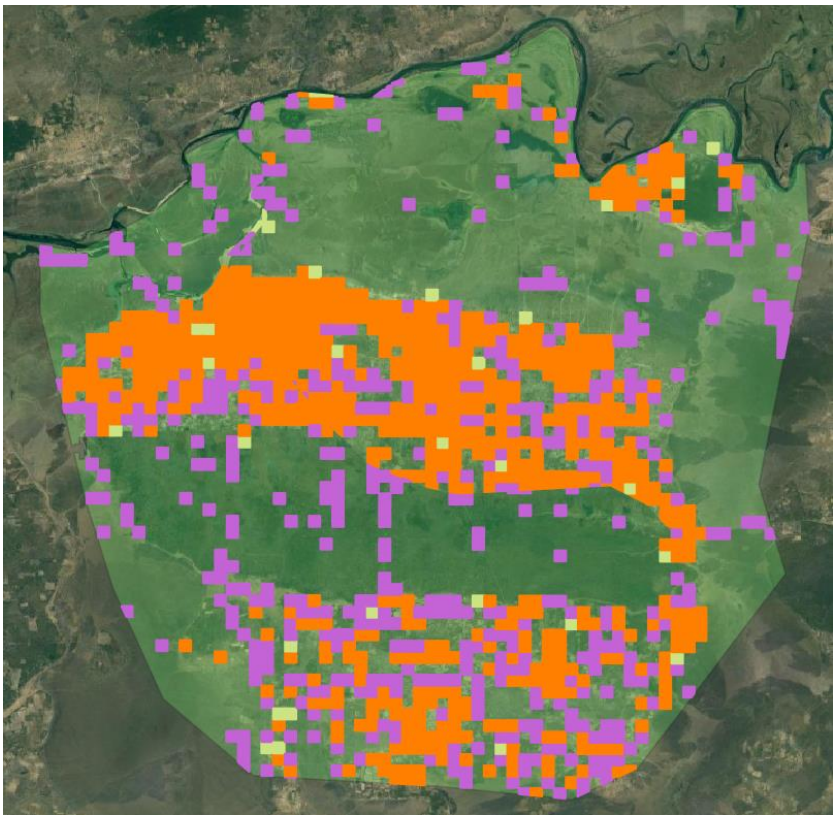
Step 4: Identifying EAs with non-zero household presence

Superimposed the roof and FB datasets onto the grid and classified each EA into one of the four categories:

1. There is at least 1 roof & FB estimates non-zero population
2. There are no roofs, but FB estimates non-zero population
3. There is at least 1 roof, but FB estimates zero population
4. There are no roofs roof & FB estimates zero population

To avoid systematic exclusion errors caused by the methods used in FB and roof datasets, the final sampling frame should contain all EAs in which there is a positive FB population projection, or in which there is at least one roof tagged (i.e. the union of EAs in 1–3 categories).

Figure 3: Nano sampling frame. Orange EAs have at least 1 roof and FB estimates non-zero population (Category 1), purple EAs have no roofs but FB estimates non-zero population (Category 2), and light green EAs have at least 1 roof and FB estimates zero population



For Nano, we had access to the locations of households from a sub-area of the chiefdom (collected during a previous round of surveying) to ground-truth this methodology and found that only 3% of category 4 EAs had at least one household present (i.e. exclusion error of 3%). In a subsequent round of data collection, we found that about 98%, 98%, and 88% contained at least one household in EA categories 1, 2 and 3, respectively.

Step 5: Determining the number of EAs in the sample

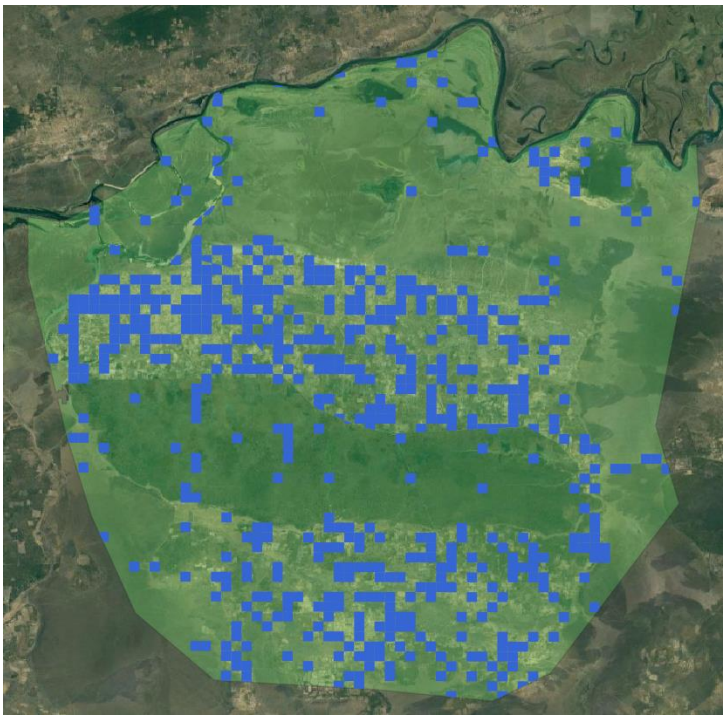
Determine the minimum number of households needed to measure the key indicators' means with the required precision, and then take a random sample of EAs that contain the minimum number of households and visit all households within the EA.

Determining the number of EAs that are needed to visit to ensure that the final sample contains the right number of households is a bit tricky and entails some guesswork. Using data from a previous round of surveying, we calculated the approximate number of households that we expected to find in each type of EA (categories 1–3), as well as the percentage of EAs that we expected to be empty. These projections were used to scale the number of EAs that we needed to sample.

Step 6: Taking a random sample of EAs

Take a simple random sample (SRS) of EAs with an equal probability of sampling. Another option is to sample proportionate to population density, which yields lower variance in the estimates (Lohr 2019)[8], Sampling Design and Analysis). However, this approach requires accurate population estimates, and ex-ante, it was unclear how well FB data would perform. Using roof density as a proxy for population size may also be problematic since higher roof density in more urban areas may be negatively correlated with the population counts (for example, market centres have a high population of roofs but a low population of households).

Figure 4: The blue cells are those included in the sample. Surveyors are instructed to visit all households within each cell (i.e. EA). This sample is illustrative and was not used for actual data collection.



Step 7: Collecting data and monitoring progress

During fieldwork, instruct surveyors to visit all households within their assigned EAs, and equip them with an electronic map. The maps should have clearly demarcated shapefiles of the EA for easy navigation. Surveyors should be instructed to visit all EAs on their maps (assuming that you only upload EAs assigned to the respective enumerator), stay within EA borders, and survey all households within the EA. In case an EA contains no households, enumerators should be required to "check-in" to the EA as proof that the EA was visited. Surveyors' progress and compliance with the procedures should be monitored remotely and as described in the enumerator management section.

[8] Lohr, S., (2019). Sampling: Design and Analysis, 2nd edition. Chapman and Hall/CRC.

IMPLEMENTATION, INSTRUCTIONS AND FILES

The code and data required to implement steps 1–5 of this strategy are found here[9] and uses Python. Users will need to download the relevant shapefiles and save them to the Shapefiles folder. The initial study area boundary shapefile should be a coordinate system that uses meters (e.g. EPSG: 32735 for Southern Africa), while the rest of the shapefiles should be in the standard longitude latitude coordinate system (EPSG: 4326).

The sample projection uses the output for the Python code and was done in Stata.[10] . Shapefiles can be loaded onto QGIS to easily visualize the data.

[9] https://github.com/IDinsight/nano-gis-sampling

[10] The Stata code is not included as this can easily be reproduced by the user.

5. ENUMERATOR MANAGEMENT

Enumerator management involves the recruitment of quality enumerators, effective training of the enumerator team, setting up adequate incentive structures, and robust data quality monitoring. This section outlines the different aspects of enumerator management that are critical for IS success.

RECRUITMENT

The term "embedded enumerators" refers to enumerators who live in the study area. Hiring embedded enumerators reduces survey costs as it eliminates the need to provide transport and accommodation to enumerators hired from elsewhere. It can also improve survey efficiency and community buy-in as these enumerators are more familiar with the region and known by communities. When working in rural areas, embedded enumerators may come with the trade-off of less survey experience and exposure to survey tools like smartphones.

When hiring embedded enumerators, use local stakeholders to shortlist potential enumerators. Local leaders like headmen, agricultural extension officers, teachers, and health workers normally work with outstanding community members who may be a good fit for the enumerator role. When consulting with these local leaders, request recommendations for potential surveyors, being as clear as possible about the purpose of the survey and minimum required competencies. The potential bias that comes with recommendations can be reduced by consulting a large number and variety of stakeholders.

Once all recommendations have been gathered, design assessment tools to screen for the best candidates. These should include literacy in English and relevant local languages, translation, arithmetic, survey comprehension tests, and in-person interviews. Include a mock survey as part of the interview to test for survey skills.

For Nano, we received 100 recommendations from stakeholders and organized assessment exercises at three venues in the chiefdom. It took approximately 1.5 hours (90 minutes) to process 30 candidates.

TRAINING

Training should be done at a centralized location with the full team to minimize effort. If required, provide accommodation for enumerators to reduce daily travel. For Nano, we found that two full days of training and three days of piloting were sufficient, and enumerators generally felt prepared on the first day of surveying.

Enumerators should be trained on research ethics and survey etiquette, use of Android devices, survey protocols, and the use of all software relevant for data collection. Software training should include data collection tools like SurveyCTO, mapping applications like Avenza maps, messaging platforms like Kaizala and WhatsApp, and online forms like Google forms. Provide a hard copy of the training manual to each enumerator to use as reference material during data collection. The manual should include guidelines on survey tools and software, survey ethics, survey protocols, emergency contacts, important definitions, and a copy of relevant survey modules.

During training, walk through the survey with the enumerators. Use this time to go through important definitions, complex questions, and survey role-plays. This is also a chance to make adjustments to survey questions and translations based on enumerator feedback.

After training, have the team pilot the survey by conducting interviews in a context similar to your study area. The pilot sample should not overlap with the final sample to avoid introducing noise in the final data. Have a debrief session after piloting and correct any issues raised during piloting. Depending on the severity of the issues, additional iterations of piloting and survey adjustments may be needed.

DATA QUALITY MONITORING

Remote supervision of surveyors includes a comprehensive evaluation of their performance with timely feedback and bonus payments to incentivize high-quality data collection.

High-frequency checks (HFC) are a standard way to track surveyors' progress and quality in real-time. HFCs run daily throughout data collection and track how many households each surveyor visits, how long each survey takes, and their compliance with protocols. Surveyors whose indicators are statistically different from the team averages are flagged and given feedback. For Nano, we coded high frequency checks in Stata and exported the outputs to an Excel dashboard. Below is a table the IDinsight Nano team received with surveyor progress. Cells highlighted in yellow signify that the surveyor average differed from the team average, which informed IDinsight on each surveyor's relative performance.

Figure 5: Snapshot of the Nano HFC dashboard

		Overall progress		Time and distance			
Enumerator	*Supervisor*	Total: HHs surveyed	Total: HHs refused or not found	Average: Consent length (minutes)	Average: Survey length (minutes)	Average km walked per day	Average length of workday (hrs)
Enumerator 1	Supervisor 1	21	6	4.35	11.73	1.81	4.26
Enumerator 2	Supervisor 1	21	6	8.36	10.36	1.35	2.86
Enumerator 3	Supervisor 1	26	1	2.94	14.34	3.18	6.05
Enumerator total		68	13				
Enumerator average		22.67	4.33	5.05	12.30	2.19	4.51

HFCs can be taken one step further by automatically sending project team leaders and supervisors messages about a surveyor's performance with instructions on how to deliver targeted feedback.

The HFC results can also be used to incentivize surveyors to follow the protocols and improve their performance with bonuses. Develop a bonus structure using a group of indicators that can be measured accurately and balance survey quantity and quality. Details on bonus payments are discussed under Salaries and bonuses.

Audio audits should be incorporated into the survey to check for adherence to survey questions and protocols. To avoid bias, employ an independent audio checker to listen to audio files. Establish a formal process for recording feedback from recordings. For Nano, we set up a Google sheet with links to audio files from the SurveyCTO server, and an audio checker listened to the audio files. For each recording heard, the audio checker recorded how closely it aligned to the exact wording of the question, whether the recorded response was correct, and the response heard from the file.

Finally, deliver "good practice" nudges to remind the surveyors about important data collection protocols. These can be delivered via automated text messages with nudges such as, "Remember to only visit households within your enumeration area" and "Read the consent form slowly and clearly."

SALARIES AND BONUSES

Salaries

Salaries should be guided by regulations and local rates, i.e. the salary should not be less than minimum government-prescribed wages while aligning with region-specific market rates. Consult with local stakeholders who work with communities (e.g. agriculture extension officers) to get region-specific rates.

Payments should account for statutory deductions like tax and pension contributions. For Nano, we collected enumerators' taxpayer-identification numbers (TPIN), social security numbers, and national ID numbers to facilitate these payments, and deducted the required amounts from the salaries.

After daily rates are set, determine the payment frequency. Salaries can be paid at the end of data collection, which is logistically easier as it is a one-off payment or paid at intervals during data collection, helping enumerators with daily cash flow. Enumerators' views on payment frequency vary, but having a payment system tailored to individual enumerators is logistically challenging. If you decide on frequent payments, each payment cycle should include payment for the number of days worked, and reimbursements for expenses incurred during that period.

In the case of Nano, we made payments twice a week to enumerators' mobile money accounts. Each payment cycle included payment for days worked and reimbursements for expenses incurred.

Bonus payments

The goal of performance-based bonuses is to incentivize effort and improvement. Design a short rubric to assign points according to signals of high effort and quality. These expectations should be communicated to enumerators before the start of fieldwork and when making payments. To encourage performance improvement, provide actionable feedback to enumerators who have not qualified for a bonus on why their performance did not qualify and necessary actions to improve performance.

For Nano, we gave daily bonuses when enumerators met a threshold of 12 points out of a possible 24, and an additional bonus for the top two performers. Bonus payments were accompanied by tailored messages on how each enumerator could improve performance.

We encountered connectivity challenges, which meant some enumerators were unable to submit forms daily. As a result, receipt of forms was often staggered, which complicated bonus point calculations. This a common problem in rural data collection and can be mitigated by providing transport for enumerators to travel daily to town to access the internet. This is important for data quality and reduces noise in bonus point calculations, improving the efficacy of the tool in incentivizing high performance.

TRANSPORTATION, COMMUNICATION, AND EQUIPMENT

A functional logistics process enables data collection to proceed with minimal disruptions. Money transfer processes should be set up to facilitate payment of salaries, transport and communication reimbursements, and other necessary logistical support.

Transportation

Assign enumerators to enumeration areas close to where they live to reduce travel time. Despite this, enumerators may still need to take a motorbike or bicycle to reach distant enumeration areas. Enumerators should be refunded for any expenses incurred to rent a bicycle or hire a motorbike taxi.

In some instances, enumerators have access to a motorbike and should be reimbursed for any fuel costs. This also entails paying for wear-and-tear and may include repair costs for any damage that occurs during data collection. If this is the case, enumerators may request advance funds to cover the costs of such repairs.

Enumerators should report expenses daily but submitting receipts or pictures of hired equipment (e.g. bicycles). Pictures of receipts or hired motorcycles/ bicycles can be submitted via WhatsApp or Kaizala. Daily expense reporting is necessary for reimbursements to be made at each payment cycle.

Communication

Enumerators should be provided with enough cell phone credit to communicate with their supervisors and purchase data for uploading forms. If possible, send airtime directly to enumerators' phones whenever required.

Provide multiple contact details for enumerators in the event of an emergency, ideally for different individuals. In our case, enumerators were given contacts for an IDinsight staff member, the MOCTA contact person, and their supervisor.

Equipment tracking

Enumerators should be provided with smartphones to use for data collection. As cell reception for different service providers may vary across the study area, provide multiple sim cards to increase the likelihood of having cell service.

Enumerators should sign a gadget release form when collecting devices for tracking purposes. The gadget release forms should stipulate possible actions if a device is lost or damaged.

REMOTE TASKING

Remote tasking entails remotely training and assigning survey jobs to enumerators. By definition, this does not require the presence of project staff in the field, reducing the cost of data collection. Remote tasking is possible when working with high-quality enumerators with survey experience or embedded enumerators who have had at least one round of in-person training. It comes with the risk of reduced data quality due to the absence of direct supervision by project staff, which can be overcome by having a trained supervisor in the field selected from the enumerator team and setting up a robust remote data quality monitoring system. The latter is described under data quality monitoring.

The following conditions have to be met for remote tasking to be possible:

1. There is a pool of enumerators who have already gone through the screening process. This should be larger than the minimum number of enumerators required to complete a survey in a stipulated timeframe. This is to make sure that enumerators declining survey tasks does not negatively impact the survey timeline.
2. Enumerators have access to smartphones or tablets, either provided by the project team or owned by the enumerators. In rural surveys, providing smartphones/tablets is necessary, as cell phone ownership is generally low.
3. Enumerators have a working cell phone and an internet connection, as this will be needed to receive survey instructions and training materials and successfully upload completed surveys.

The first step is to check on enumerator availability by sending a message (either via SMS or other messaging platforms) detailing survey duration, start and end dates, and requesting their availability.

Once enumerators confirm availability, send them training materials (in pdf format) via WhatsApp, Kaizala, or alternative messaging platforms. Alternatively, you can send training materials to the supervisor and ask him/ her to print and share with enumerators who have confirmed availability.

After the enumerators have accessed and read the training materials, ask them to complete a quiz via an online survey platform (e.g. Google forms). The quiz should include questions on survey content, protocols, and procedures described in the manual. Multiple choice questions are best suited for the quiz as they require little to no processing. Each quiz question should be weighted based on the number of possible correct answers for that question. For example, select-multiple questions (multiple correct answers) should be assigned more points than select-one (single correct answer) questions. A pass threshold should be established and clearly communicated to enumerators. Set and communicate the deadline for completing the quiz and the number of allowed re-attempts by each enumerator.

When the minimum number of required enumerators complete the quiz, send out instructions on how to access the survey on SurveyCTO, enumeration area assignments, and digital maps.

6. DATA PRESENTATION

Data should be presented in a simple way that allows users to diagnose problems and develop context-specific solutions. This is especially important when working with users with little to no background in statistics.

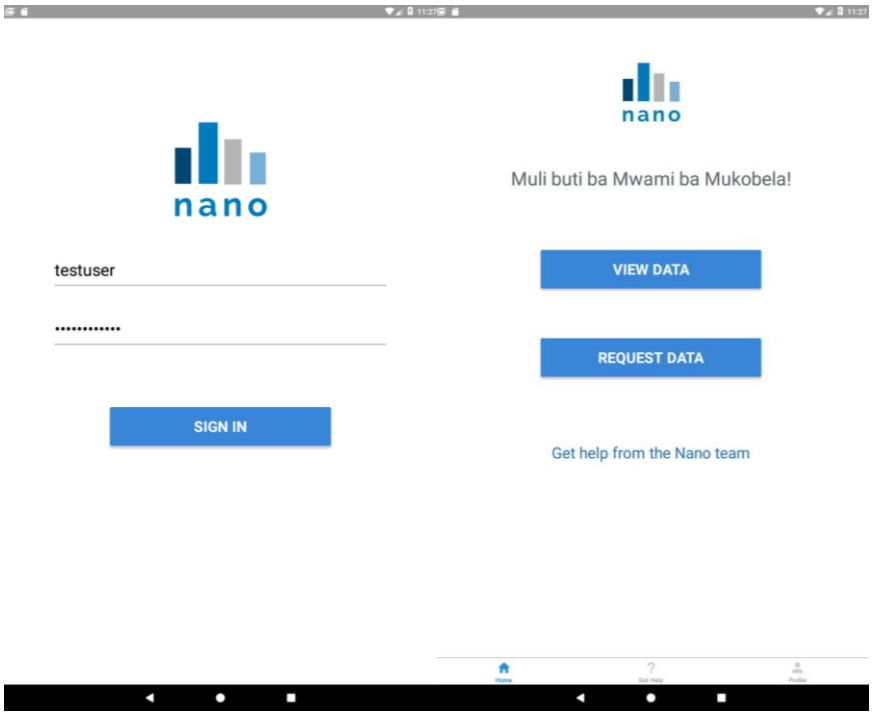
The first step in creating an intuitive and informative display is selecting the correct visualization. Present prototypes and ask the user to choose which visualization option are most intuitive, and then incorporate the feedback into the data presentation.

The next step is to select the platform(s) on which to present the data. Options include paper reports, web-based dashboards, and mobile applications. Applications and dashboards can be easily updated as new data is collected, and allow for a broader range of visualizations compared to reports. However, they require a high level of technical capacity to design and build, which may have to be outsourced.

The final step is to train the chief to use the selected platform, read and interpreted the data. If a digital platform was selected, the user should be trained to use the tablet/computer on which the data is presented, and how to navigate the dashboard or application.

We developed an Android application to present data and nudge the chief towards actions. The app is password protected to limit access to the data.

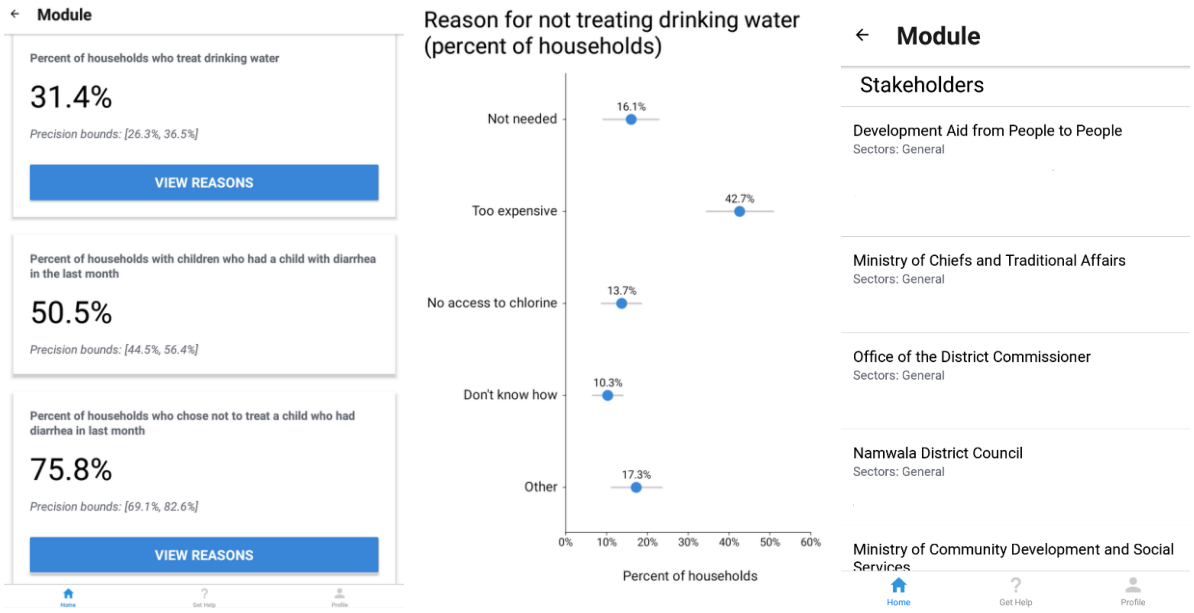
Figure 6:Nano app welcome screen



The app uses tow visualization options; cards with point estimates and confidence intervals in brackets and, chart with point estimates as dots and confidence intervals as error bars. These options were

selected based on feedback from the chief. The Nano app also includes contact information for stakeholders at the bottom of each screen, facilitating communication and increasing the likelihood that the chief will act on the data. At his request, we also produced a paper report that presented data in table and graph formats. Producing paper reports was useful for sharing data with stakeholders, but may be difficult to do at scale.

Figure 7: Nano app data visualization options[11]



[11] Individuals names and contact details are not shown to maintain privacy.

7. CONCLUSION

The results from the Nano Pilot have been positive. We successfully recruited and trained embedded enumerators, completed two rounds of data collection, developed and android applications and, presented data to Chief Mukobela. The chief has positively engaged with the IS, and the data presented to him has changed his mind about the chiefdom's needs and challenges.

This guide provides guiding principles and actionable advice on how to replicate Nano, which can be applied to different contexts and levels of governance. Adhering to the principles outlined in this guide, we hope readers can successfully replicate parts or all of Nano and provide decision-relevant data to users cheaply and quickly.

8. APPENDIX

1. Sample interview guide

Qualitative Interview: Chief /Senior Headmen/Ministries

Metadata Interviewer should fill this in after consent is given and before interview begins.	
Interviewer name	
Interview Date and Start Time	D D / M M / Y Y
	H H : M M
Chiefdom	
Name of respondent	
Title of Respondent	
Mobile number of respondent (if applicable)	

Note: This part should include instructions on the purpose of the interview, interview structure and any additional instructions to the interviewer.

Section 1: Mechanics/Design **Only Applicable for Chief and Senior Headmen who actively use the Nano dashboard**	
1. Were you able to find the information you were looking for on Nano? **If yes**, how easy or difficult was it to find this information?	
Section 2: Data Use	
Module-specific questions: For each module ask the following questions	
Trust in the data: Did you believe data in module X is correct? **If no,** what do you think the data should be showing instead?	

Alignment with the priors:	
1. In Module X, which of the statistics did not surprise you i.e. showed what you expected to see? 2. In Module X, which of the statistics surprised you? a. Why did they surprise you? 3. Did you learn anything about your Chiefdom? If so, what did you learn?	
Section 3: Customer Satisfaction - "Love" For the Product	
1. Would you use Nano to request data collection again in the future? **If yes,** what kind of areas/topics would you want to collect data on? **If no,** why would you not use Nano again?	

2. Example module assessment matrix

Survey module	Interest from Chief	Interest from others	Decision relevance	Feasibility of quality data collection	Training ease	Involves other ministries	Overall assessment
Module 1	Very high	Very high	Medium	Medium	Medium	Yes	Implement
Module 2	High	Medium	High	High	Medium	No	Implement
Module 3	High	Medium	High	High	Medium	Yes	Implement
Module 4	Medium	Low	Medium	High	High	No	Do not implement
Module 5	Medium	Medium	High	High	Medium	Yes	Implement
Module 6	Medium	Medium	High	High	High	No	Implement
Module 7	High	High	High	Low	Low	Yes	Do not implement
Module 8	High	Medium	Medium	High	High	Yes	Implement

Note: In assessing the viability and value of modules, considered several factors that could affect the eventual social impact of the data. The initial list of potential modules should be constructed solely based on interest from the chief, as his engagement is the primary factor in determining social impact.

Additionally, consider interest from other stakeholders, particularly ministry counterparts, because action in this context typically requires collaboration across ministries. similarly, prioritise modules in which you expect the chief to have the authority and capacity to take meaningful action. Favour modules for which you expect to feasibly collect high-quality data and train inexperienced data collectors to measure properly.