Developing a Lean Data Management System for an Emerging Social Enterprise

Eric Obeysekare, Anthony Marucci, Khanjan Mehta
Humanitarian Engineering and Social Entrepreneurship (HESE) Program
School of Engineering Design, Technology and Professional Programs (SEDTAPP)
The Pennsylvania State University
University Park, PA
Correspondence: khanjan@engr.psu.edu

***Abstract*— Data collection is essential for organizations operating in developing countries and other low-resource environments. Whether for traditional monitoring and evaluation, efficiency improvements, or research purposes, data collection is an essential part of the scaling process of any venture. Despite the necessity of data collection, creating an efficient system that captures key information is no easy task. Choosing the proper tools, designing comprehensive protocols, and integrating data collection into organization workflows are essential components of any data management system. To accomplish this, organizations are turning to a Lean approach of reducing all possible inefficiencies, which despite its origins in the manufacturing field, is becoming popular among entrepreneurs and researchers alike. From a data collection perspective, eliminating waste, such as inefficient forms or collection of un-needed data, allows organizations to create efficient systems that serve their purposes and provide needed information in a timely fashion. This article will describe the data collection system being implemented by GRO Greenhouses, a social enterprise that sells greenhouses in several countries in sub-Saharan Africa. Using off-the-shelf tools like Google Docs, GRO was able to create an easy-to-use system that addressed the organization's needs as it expanded from its start as a university-based research project into a full-fledged venture operating with external funding. Collecting relevant data is the next step in GRO evolution as long-term sustainability becomes more important. Lessons learned by GRO about data collection and scaling are relevant to other similar ventures as they seek to evolve and expand.**

Keywords—lean data collection; social enterprise; developing countries;

I. INTRODUCTION

The international development community has recognized the benefits that come from collecting high quality datasets. Data that express the current state of operations can be a valuable resource for non-profit and for-profit organizations working in developing countries, contexts where limited resources must be put to the best possible use. Better decisions about operations can be made with the aid of data leading to efficient operations and improved outcomes. Given that many development projects function using external funding, data is also needed for reports to be sent back to those funders. Accountability can also be improved by better tracking projects as they evolve – another area where data are essential.

However, the focus in recent years has been on Big Data, the practice of incorporating large-scale dataset analysis into decision making, with some experts calling for the use of Big Data for Development [1] and others recognizing that Big Data can be a useful tool in order to make development policy decisions [2]. However, while Big Data is indeed a promising emerging field, it is not always the best solution for international development projects as they attempt to scale from initial funding into long-term sustainability. These types of organizations need to prioritize data collection in a manner that provides necessary amounts of data that can be efficiently managed and effectively harnessed during decision making. Datasets can quickly become overwhelming as they grow in size; something that early stage ventures should consider when deploying their limited resources to the task of data collection. Instead, a focus of collecting and analyzing only the data that are necessary, taking inspiration from the lean philosophy of eliminating waste wherever possible, is necessary. This type of 'lean data' approach is being advocated for as a better choice for businesses which often do not need the massive datasets called for by the Big Data movement [3]. These ideas have also spread to the field of international development where lean research is being championed as a promising strategy for those conducting research in developing countries [4]. When local communities are repeatedly recruited for long surveys year after year, frustration can set in among residents. Lean research asks that researchers keep this in mind when conducting studies and pare data collection down to only the essential questions that point to specific outcomes. Social enterprises in particular can turn to lean data as they attempt to become profitable and need actionable data which can be collected efficiently [5]. A lean data strategy emphasizes flexibility, adaptability, efficiency, and importantly, impactful data. Social enterprises, especially those in the early venture/start-up phase, must be able to quickly pivot data collection efforts in order to efficiently collect data that can directly lead to decisions about resource deployment, project evaluation, and organizational improvements that advance the venture's economic and social bottom lines. One such social enterprise is GRO Greenhouses, a venture attempting to evolve into long-term sustainability.

GRO Greenhouses is a non-profit social enterprise selling affordable greenhouses in Sierra Leone and Mozambique. Originally started as a collaboration between World Hope International, an NGO operating in several developing countries including Sierra Leone and Mozambique, and the Humanitarian Engineering and Social Entrepreneurship Program at the Pennsylvania State University, GRO sells affordable greenhouses that farmers receive on a cost-recovery

basis and work to pay off over time using profits from sales. GRO also offers training and agricultural extension services to its customers.

As GRO scales up operations, a need to create a lean data management system was recognized. GRO needed a system that easily allowed for reports of important aspects of the data whenever they were needed. Since GRO operates in multiple countries, the system needed to work in multiple contexts, pointing to the need for a centralized, possibly cloud-based solution.

This article presents a case study of the lean data management system designed and implemented by GRO Greenhouses. After reviewing related work that addresses data collection by humanitarian organizations, an introduction to the operations of GRO is given. Next the selection criteria, alternative system options, and design decisions that were made about the final data management system are presented. The data collection and analysis workflow is then presented including specific examples of each step. Finally, some remarks are made about the implementation of lean data management system by humanitarian organizations. This article is of interest to ventures that are attempting to scale up from small projects, research or otherwise, into profitable social enterprises. The authors hope that anyone reading this article will be able to utilize the insights presented here to construct their own lean data management system that fits the specific needs of their organization.

II. Related Work

Data collection is a task that many humanitarian organizations must undertake in order to effectively monitor and evaluate their projects. This has resulted in the development of several data collection systems designed specifically for use by humanitarian organizations operating in low-resource contexts. These systems can be divided into three categories: mobile data collection systems, sensor monitoring systems, and computer vision based systems that complement existing data collection operations.

Perhaps the most well-known of the mobile data collection systems is Open Data Kit (ODK) [6]. ODK provides a centralized server to which data can be aggregated from mobile phones, either through voice prompts or the use of customizable fillable forms. ODK has inspired other similar data collection systems including KoBoToolbox, Ona, OpenSRP, MDC, ArcGIS Collect, Data Winners, and mField Work. These other systems are further discussed in the system selection and evaluation section.

The ODK system has also been adapted in order to collect data from sensors that can be deployed in the field to collect more specific, technical data [7]. Other data collection systems have been created to integrate directly into existing workflows in low-resource contexts which usually rely on paper as the main medium of data collection. mScan [8] and Shreddr [9] are two such applications that use computer vision technology to automatically digitize paper forms into a centralized database.

Data analysis is another important part of data management. All of the systems mentioned here have basic data analysis capabilities with some of them offering more advanced features like data visualization (charts and graphs) or mapping. However, none of these projects offer the flexibility and ease of development and use that is essential for early stage social entrepreneurial ventures.

III. GRO Greenhouses Introduction

GRO Greenhouses is a social enterprise operating in Sierra Leone and Mozambique. GRO sells affordable greenhouse constructed with local materials. Using one of these greenhouses, farmers are able to complete shorter crop cycles year round using less water while also improving yields. GRO greenhouses protect plants from extreme weather while also creating a microclimate inside the greenhouse that facilitates plant growth and water retention. The design used by GRO has been tested in several different contexts, each with its own unique climactic conditions and supply chain scenarios, and is currently commercialized in Kenya, Tanzania, Cameroon, Mozambique, and Sierra Leone.

Now in its third year of existence, the GRO Greenhouses venture has been steadily growing operations through increased sales, in addition to receiving several grants. With an initially small of number of greenhouses, it was possible for GRO staff to keep track of the greenhouses informally via conversations, spreadsheets, and email. However, as sales increased, these informal data management systems quickly became impractical and unsustainable.

Data management is needed by GRO Greenhouses in order to monitor each greenhouse, provide agricultural extension services like agronomic support and market linkages to farmers as needed, and report metrics and results to funding sources. GRO sells greenhouses on a cost recovery basis, allowing farmers to pay back the cost of the greenhouse over time as vegetables grown in the greenhouse are sold. Tracking the crops that are grown in each greenhouse allows GRO staff to evaluate the relative success of each greenhouse and determine a repayment structure and schedule that can feasibly be met by the farmers. Greenhouse tracking and general data management allows GRO to efficiently deploy limited personnel resources to provide extension services. Finally, GRO must report progress towards several metrics put in place by funding providers. These metrics are aggregated totals of various venture outputs including weight and cost of produce sold, greenhouses constructed, and volume of water saved, among other things. Aggregating and calculating these totals requires a centralized location where data about all GRO greenhouses can be stored and analyzed.

These requirements point to the need for a comprehensive digital data management system that can be used by GRO staff and also further scale as GRO continues to expand operations. This system will be used to supplement the paper records that are traditionally kept by an organization like GRO. Paper records present several workflow challenges as it is difficult to update records, aggregate information, and synchronize data across many countries using only paper forms. Since GRO Greenhouses operates in many countries, and needed an easier way to aggregate data, a digital system was needed.

Prior to beginning the design and development of the system, an evaluation of existing data collection and management platforms was conducted including both existing, off-the-shelf systems and the possibility of developing a new system from scratch. Each option was ranked based on several criteria, which will be described in the following section.

Following this comprehensive evaluation, a decision was made and design and development of the system was conducted.

IV. System Selection and Evaluation

A. Selection Criteria

When evaluating the possibilities for a data management system, the following criteria were used: cost, ease of use, ability to make custom reports, need for server hosting, and ability to enter data on a desktop computer. Since GRO is a small non-profit, paying for a data management system was not an option. The system needed to be easy to use so that any employees could quickly pick up the system and begin to use it. Custom reports were required so that GRO could keep digital records of all its greenhouses. Hosting needed to be provided for the system to be easy to maintain and to keep costs down. Finally, most people in Sierra Leone and Mozambique do not want to use the internet on their cell phones for work because of the high cost. The system therefore needed the ability to input data from a desktop computer in the office where there is an existing internet connection.

B. Existing Off-the-Shelf Solutions

There are many existing data collection systems. Most existing systems focus on data collection, not management, and do not allow for easy aggregation and analysis of data. They are designed for mobile data collection, something that was not feasible for GRO Greenhouses.

1) Paid Options

ArcGIS, DataWinners, and mField Work are three examples of data collection systems which have a fee associated with use. ArcGIS, for example, has many different plan levels with the least expensive being for only 5 users and costing $2500 a year. GRO would need more than 5 users, and the next option up is for 50 users totaling $10,000 a year. ArcGIS also is more focused on mapping and data collection as it relates to those mapped locations. DataWinners, another data management system, has a free plan but only allows for 1000 submissions per year. With data collected weekly for many greenhouses, this number would quickly be surpassed. The next cheapest option for DataWinners is $99 a month for a full year and $199 a month on a month-by-month basis. mFieldWork is the arguably the most affordable option with a pricing model based on submissions, and costs $0.30 a submission. mFieldWork, however, is more targeted for medical work. Ultimately, paying for a data collection system was not an option for GRO and the options presented here were discarded.

2) ODK

Open Data Kit (ODK) is a popular solution for data management. It is free and open source. However, with ODK hosting can be an issue. ODK systems can be hosted on Google App Engine, a free option but with limited bandwidth, or on a personal server that can be costly. Additionally, with ODK, there is no easy way to enter data from a desktop computer; data entry is intended to be done using a mobile phone or tablet. There are tools that allow for desktop entry, but they require hosting and hence with ODK, there is no way to have desktop data entry without purchasing hosting.

3) KoBoToolbox

KoBoToolbox is another system designed for NGOs to use for data collection. However, KoBoToolbox only allows for simple reports. It does not allow for custom queries, which GRO Greenhouses needed.

4) Ona

Ona is very similar to KoBoToolbox. It is free and provides hosting, however it does not allow for custom queries. GRO Greenhouses needed to be able to implement custom queries for its data management system.

5) OpenSRP

OpenSRP is an open source mobile health platform. GRO Greenhouses considered trying to use this, however hosting was not provided, a mobile phone must be used for data entry, and it does not allow for custom queries.

Table 1: Data management system decision matrix

	Desktop Entry	Remote Changes	Ease of use	Pictures	Multiple forms	Users	Server provided?	Free?	Database Access
MDC	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
ArcGIS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Open Data Kit + Enketo	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
KoBoToolbox	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Ona	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Data Winners	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
OpenSRP	No	No	Yes	Yes	Yes	Yes	No	Yes	No
mField work	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Google Forms based system	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

6) MDC

MDC is a mobile data collection system. It does not allow for custom queries, and it also does not allow for desktop entry.

C. Custom System Design and Development

An entirely custom system would meet all of the criteria. However, a custom solution is much harder to design. It would require hosting as well as a domain name. A database would have to be designed and installed onto the host. From there, HTML and PHP pages (for example) would have to be designed for each form, which would require a not insignificant amount of programming. Finally, all of this would have to be styled with CSS so that the forms and outputs are more readable and able to be understood by everyone. While templates exist for forms, they still require some amount of programming knowledge to function with PHP and the database. Lifecycle issues were also a concern with a custom-built system that would require additional resources to maintain, update, and further grow over time. Ultimately, the resources required for a custom designed and built data management system were not available for a venture like GRO Greenhouses, which is currently focusing resources on scaling efforts.

D. Decision Matrix

The majority of the systems mentioned here did not fit the criteria of desktop entry, easy to use, allowed for pictures, allowed multiple forms, hosting provided, free, and allowed for custom queries. Google Forms and Sheets, however, allowed for all of these and was the system chosen for use in GRO's data management system. Table 1 shows a decision matrix that outlines the features of each of the systems mentioned here.

V. Design Decisions

When designing a data management system for an organization operating in low-resource contexts, there are many factors that must be taken into account including the existing organizational workflow, technical infrastructure that may or may not exist (e.g. device availability, internet connectivity, etc.), and technological literacy. Specific factors that were considered when designing the GRO data management system are discussed in this section.

A. Paper and Digital Forms

The system needed a way to collect data without changing existing workflows. Collecting data on paper was already part of the existing GRO workflow. This is common in developing countries where paper is often the first choice for data collection and storage over electronic options. This is often due to under developed technical skills and infrastructure that can support digital data collection, ultimately making paper data collection the most straightforward option. While smartphones are becoming more common in the developing world, mobile internet access is often quite expensive. In the case of GRO, greenhouses are often constructed in rural communities that are not well-connected. It is also not practical to bring a laptop for fieldwork, when the data can be directly collected using paper forms.

For these reasons, it was determined that a data management system for GRO should consist of two parts: paper worksheets and corresponding digital forms. The paper worksheets could be easily used in the field and fit into the way that GRO staff currently worked. Data would then be transferred to the digital forms for storage and analysis.

B. Simple, Intuitive Design

To ensure that new employees are quickly and efficiently integrated into GRO workflows, a simple, relatively non-technical system was designed. Using spreadsheets, which staff are already familiar with, to store the data made the system easier to understand. In addition, the actual data collection forms were designed to be intuitive and straightforward. They also include step-by-step instructions that specify exactly what data need to be collected. This lean approach ensures that no unneeded data will be collected.

GRO also operates in multiple countries where English can be considered a second or even third language. Keeping the system simple and straightforward ensures that staff in each country will have the same interpretation of data collection. With GRO's limited resources, it was not feasible to create different interfaces for each country meaning that the system truly needed to be comprehensible by all.

C. Off-the-shelf

A solution was needed that did not require a lot of programming. GRO Greenhouses did not want to expend resources on a system that might not fit its needs. Using off-the-shelf software is advantageous in other ways. Maintenance of a data collection system is something that organizations should consider. Custom-created solutions must be maintained internally, while off-the-shelf products are maintained by their creators. In the case of Google Forms/Sheets, GRO can trust that these products will be maintained and advanced regularly. This will allow the data management system to evolve over time in order to take advantage of advances in the under-lying technology.

D. Work on Laptop or Desktop

Many individuals in sub-Saharan Africa do not want to use their personal cell phones to work on the internet because it is very expensive. GRO Greenhouses, however, provides internet access at the office for its employees. This internet connection can be used to digitize data and the use of a laptop or desktop may make this easier than using a mobile phone. Therefore, it was not feasible to have a system that would only allow for mobile data entry. This system using Google Forms allows GRO employees to enter data from a laptop or desktop computer, and from mobile phones or tablets should future workflows call for that ability.

E. Data Aggregation and Reporting

Since multiple forms were designed to simplify the process for GRO employees, a system was needed that could combine the results from many forms into one report. GRO employees enter data in discrete forms, and the system is able to query each of those and combine the results to analyze the data. Using this is it then possible to filter the results of all forms by greenhouse number to see the output of everything for a given greenhouse.

F. Greenhouse Number System

A logical numbering system for greenhouses had to be designed. Since GRO greenhouses is in multiple countries, the

number involves a country code. Each greenhouse number starts with the abbreviation for the country it is in. This is SL in Sierra Leone, and MZ in Mozambique, the two locations where GRO is current operating. The next part is a number indicating the order the greenhouse was built. Finally, a forward slash followed by the year the greenhouse was built. So for the 10th greenhouse built in Sierra Leone in 2015 would have the number SL10/15. These greenhouse identification numbers help GRO to keep track of each individual greenhouse. Each form that is filled out by GRO staff includes one of these greenhouse numbers and archived forms are organized by greenhouse number. This asset-tracking style approach again ensures that only necessary data is collected.

VI. GRO Data Management Solution

GRO Greenhouse has implemented a system that provides a complete data management solution including data input, storage, analysis, and reporting. It was designed using Google Documents off-the-shelf software and is entirely in the cloud. The system was designed with flexibility in mind and is able to be extended to many purposes. This section will describe the system in further detail.

A. GRO Data Management Workflow

Data collection is an ongoing process that happens at each different stage of a greenhouse's lifecycle: planning and construction, actual use growing vegetables, and ongoing maintenance procedures. shows a process diagram which details GRO's data collection. The process diagram is broken up into four vertical columns that show who/what is needed for each activity including Farmers, GRO Staff, Paper Forms, and Digital Data Entry. This diagram was created following a systemic analysis of GRO operations that distilled the everyday workflow of staff into discrete components that fit well as data collection tasks. This analysis was particularly important while designing a truly lean data system. Each field that was identified during this analysis was examined in order to determine if it was truly necessary to collect.

The Farmer and GRO Staff columns show the interactions that occur between the two parties. Data collection goes along with each different type of interaction and a corresponding paper worksheet exists for each (the Paper Form column of the process diagram shows these worksheets). Data from these forms will also be entered digitally into the matching online form (the Digital Data Entry column).

B. Data Input

To input data into the data collection system, a combination of paper and digital forms were created. Paper forms are used when data is being collected in the field. Data from these paper forms are then copied to the corresponding digital forms upon returning from the field. Paper forms are then archived in the GRO Greenhouses archive and are available for review if needed (in the case of an audit by funders, for example). This archive consists of a file cabinet system that contains a folder for each greenhouse, labeled by greenhouse number. Looking at the folder for a given greenhouse would allow for all paper worksheets associated with the greenhouse to be reviewed showing its comprehensive life history.

GRO Greenhouses currently uses seven different forms to track and monitor greenhouses. They include the Scoring

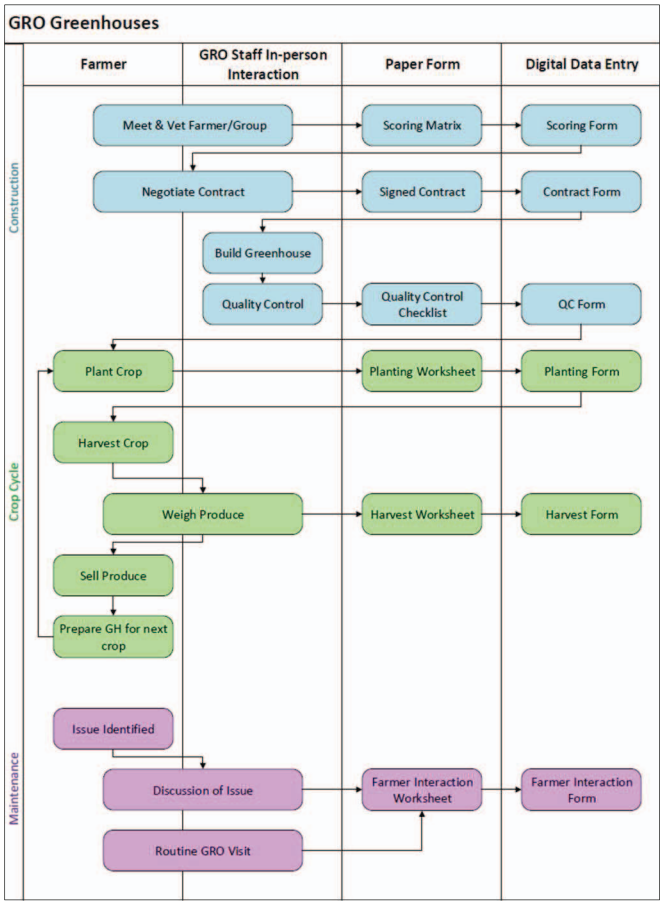
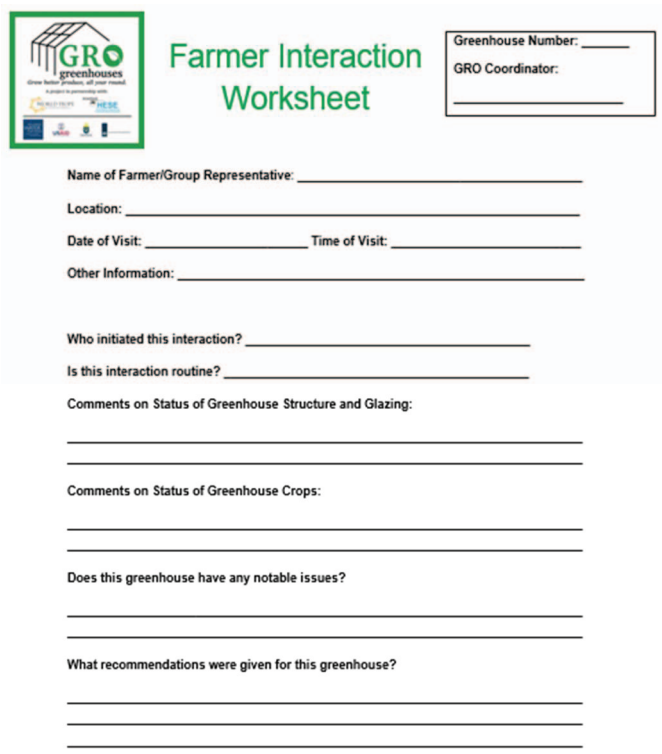


Figure 1: GRO Greenhouses Process Diagram

Matrix, the Contract Form, the Quality Control Form, the Planting Form, the Farmer Interaction Form, the Harvest Form, and the Payment Form. Together these forms collect data about all necessary aspects of the greenhouses. Figure 2 shows an example of one of the paper forms used by GRO: the Farmer Interaction Worksheet.

Each of the paper worksheets and corresponding digital forms is designed to capture data about a specific portion of the greenhouse's lifecycle. The Scoring Matrix is used to determine if a candidate is qualified for a cost recovery based sale and is evaluates that individual's potential for success with a greenhouse. The Contract Form is a form that inputs all the information for the contract. With this it is possible to see the exact terms every greenhouse has and have record of the contract. The Quality Control Form is used to make sure that each greenhouse meets standards established by GRO. This is kept on file so that there is proof it was constructed correctly if there were to later be a dispute. The Planting Form is used each time a farmer plants a crop and keeps track of the type of crops planted, how many were planted, etc. The Farmer Interaction form (Figure 3) records all information when a GRO staff member visits a farmer. The Harvest Form records all information about each harvest. It has fields for what kind of crop, how much was sold, how much was grown in total, etc. The Payment Form keeps track of all payments that were made. Combining this with the Contract From it is possible to see when a farmer will finish paying off a greenhouse, and how much they have left to pay.



Farmer Interaction Worksheet

Greenhouse Number: _____

GRO Coordinator: _____

Name of Farmer/Group Representative: _____

Location: _____

Date of Visit: _____ Time of Visit: _____

Other Information: _____

Who initiated this interaction? _____

Is this interaction routine? _____

Comments on Status of Greenhouse Structure and Glazing:

_____

_____

Comments on Status of Greenhouse Crops:

_____

_____

Does this greenhouse have any notable issues?

_____

_____

What recommendations were given for this greenhouse?

_____

_____

_____

Figure 2: Farmer Interaction Worksheet

GRO extension agents use these paper worksheets to collect data in the field. The paper worksheets have the same questions in the same order as the online forms, which were designed using Google Forms. Google Forms allows for different types of data to be input for each question, including numbers, text responses, location, and date. When the GRO staff member returns to the office from the field, they input all data collected on paper worksheets into the digital Google Forms. Figure 3 shows an example of one of these digital forms, the Farmer Interaction Form which corresponds with the Farmer Interaction Worksheet shown in Figure 2 above.

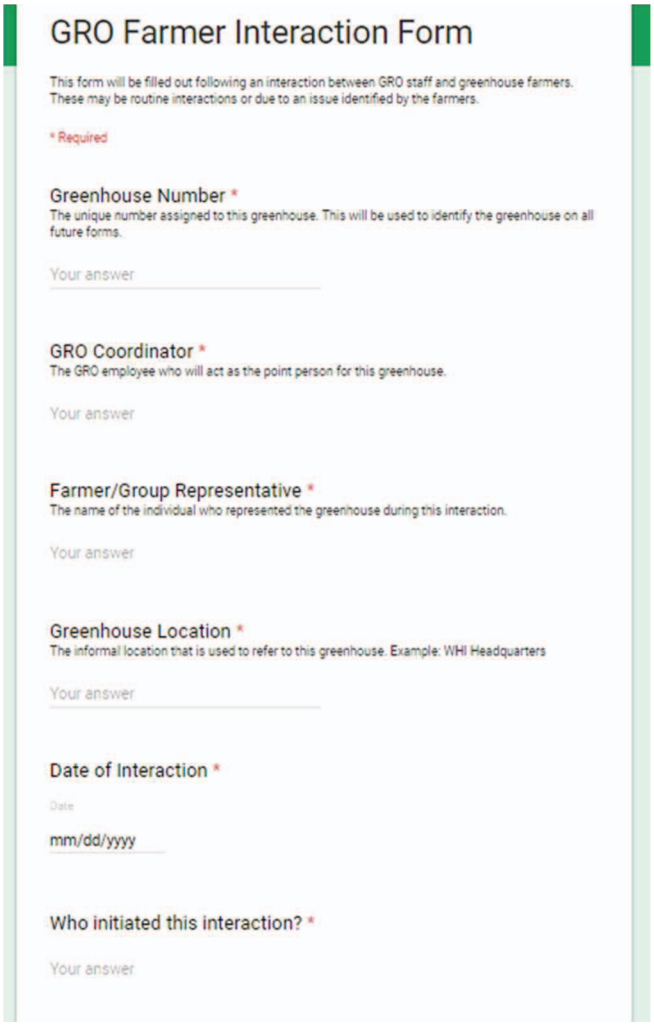
When one of the digital forms is submitted, the data are recorded in a corresponding Google Sheets spreadsheet. The spreadsheets can also be updated manually to correct any data entry errors or update inputted data as necessary. These spreadsheets can then be used to analyze and output the data and create reports.

C. Data Outputs

Data from the GRO Greenhouse data management system can be accessed in two ways: by viewing raw data as it was directly input into the Google Forms or by accessing custom created reporting spreadsheets. Both of these methods are accessible via the cloud, enabling GRO staff to access data stored in the system from any location, something that is necessary for organizations that operate in multiple locations

1) Raw Data

With this system, it is possible to see the raw data allowing for easy viewing of all information. For example, it is possible to see all payments that have been made for each greenhouse by looking at the raw input of the payment form. This also



GRO Farmer Interaction Form

This form will be filled out following an interaction between GRO staff and greenhouse farmers. These may be routine interactions or due to an issue identified by the farmers.

* Required

Greenhouse Number *
The unique number assigned to this greenhouse. This will be used to identify the greenhouse on all future forms.

Your answer

GRO Coordinator *
The GRO employee who will act as the point person for this greenhouse.

Your answer

Farmer/Group Representative *
The name of the individual who represented the greenhouse during this interaction.

Your answer

Greenhouse Location *
The informal location that is used to refer to this greenhouse. Example: WHI Headquarters

Your answer

Date of Interaction *

Date

mm/dd/yyyy

Who initiated this interaction? *

Your answer

Figure 3: Farmer Interaction Form

allows for comparisons between multiple greenhouses. However, this method may by confusing as it contains information from all GRO operations, spanning multiple countries and years. Nonetheless, having access to the data in their raw form may occasionally be useful. Data quality can be spot-checked and mistakes that may have been made during data entry can be corrected. It is anticipated that accessing the raw data will happen only rarely in favor of custom reports that provide a more comprehensible view of the same data.

2) Custom Report Spreadsheets

For reporting purposes, several report spreadsheets were designed. Given the flexible nature of the data management system, additional reports can be created on-demand as needed. The GRO Greenhouse data management system currently has two report spreadsheets that can be used to view data about the status of greenhouses: a general greenhouse information spreadsheet and a financial report spreadsheet

a) General Greenhouse Report

The general greenhouse report is a Google Sheet that displays all information about any given greenhouse by querying multiple sheets to find information about payments, crop cycles, and extension worker visits. A greenhouse number can be entered into a cell at the top of this spreadsheet. The

system then uses the SQL derived Google Query Language [10] to aggregate data corresponding to that greenhouse from the raw data spreadsheets that have been populated through the digital data input forms.

Using the general greenhouse report, it is possible to see an individual greenhouse's life history. This includes data from the Planting Form in order to track each crop cycle that was planted at a given greenhouse. Harvest data corresponding to that planting is also displayed. Additionally, each time a GRO staff member interacts with farmers from that greenhouse, it is recorded via the Farmer Interaction Form and displayed on this report. If any problems arise with a greenhouse, the history of that greenhouse can be quickly brought up allowing GRO to evaluate where things went wrong. Using this chronological report of each greenhouse's activities, GRO can also evaluate where resources like extension workers are being used and where they should be deployed in the future.

Figure 4 shows the general greenhouse report below. After the user enters a greenhouse number in the yellow box, the system will query each of the data storage spreadsheets and display a summary of the activities that have been conducted with that greenhouse. Using the other tabs in the spreadsheet (shown at bottom of screenshot), the details of specific interactions can be viewed. For example, by navigating to the Planting tab, information about each time a crop cycle was planted in the greenhouse is displayed.

Greenhouse number	SL16/15		
Information:			
GRO Coordinator	Farmer/Group Representative	Village	Chiefdom
Musa Tholley	Daniel D. Thullah	Mayawlaw	Pakie-Masabong
Remaining Balance:	2050000		
History of Preformed Actions:			
Date	Action		
1/7/2016	Payment		
3/29/2016	Payment		
4/26/2016	Farmer Interaction		
5/9/2016	Farmer Interaction		
5/25/2016	Farmer Interaction		

History by GH# | Payment | Planting | Harvest | Farmer Interactions

Figure 4: General Greenhouse Report

b) Financial Report

The second report currently being used by GRO is a financial report that tracks the economic output of each greenhouse, a metric that GRO is interested in tracking as sales increase. Using data from the Harvest Form (which collects information about the amount of produce from the harvest that was sold), this report calculates the expected profit from a crop cycle. In order to do this, the current market price and a greenhouse number are entered into a cell at the top of the report. The system then pulls data from the Harvest Form raw data spreadsheet and calculates the market value of that harvest. Combining this information with the Payment and Contract Forms, it is also possible to track the repayment of those greenhouse that were sold on a cost recovery basis. Using this report enables GRO to stay up-to-date on the farmers' finances and determine the best repayment plan for each greenhouse ensuring that there are no financial surprises for either GRO or the greenhouse farmers.

VII. Recommendations for Design

The Google Documents-based lean data management system has been implemented by GRO Greenhouses and is currently being used in both Mozambique and Sierra Leone. With operations in multiple locations, creating a system that was easily accessible from anywhere was important for GRO. Furthermore, GRO staff already had some experience with Google Documents making the platform a good choice for GRO's data management needs. Selecting Google Sheets to create reports also allowed for the use of Google Query Language, a robust query language that was able to accomplish most of the required data aggregation and analysis in a straightforward fashion. Using Google Query Language, GRO will also be able to expand existing reports and create new reports as the need arises.

GRO's decision to use both paper and digital forms has proved to be a good choice. Paper forms are easily able to be used during fieldwork, which is where GRO staff most often collect data, and also provide a hard record of activities conducted should the need for review arise. Creating a systematic archive for these completed paper forms has also proved successful.

While GRO's implementation of a lean data management system has been mostly successful, there are still some caveats and limitations to the system. Data quality can be difficult to verify as those reviewing data for GRO are for the most part based out of country. Audits can be conducted on site to compare the archived paper forms with data entered into the system but this can be time consuming and will only happen occasionally. One particular challenge in this arena is the use of the greenhouse number system. A specific format was designed for these greenhouse numbers (see details above) but the system depends on exact consistency when entering this number. If an extra space is inserted into the number where there should be none, the system will have trouble finding that data when creating reports. However, this is a workflow issue and will be corrected through the creation of a paper table that shows every greenhouse and its corresponding number. Internet connectivity can also be somewhat of an issue for this system as it is based in the cloud and requires an internet connection to access and enter data. However, while internet outages in the locations where GRO operates are not uncommon, they have not had a negative effect on this system. The paper forms serve as a fail-safe during internet outages and allow GRO staff to enter data as is convenient for them. The nature of the data being collected (mostly textual information) do not require large amounts of bandwidth to upload and as such do not require an internet connection with a large amount of bandwidth.

GRO has also benefited from the implementation of several other informational tools including a chalkboard tracking system and an intra-team messaging group based on WhatsApp. The chalkboard, installed in the main GRO office, is used to keep track of every greenhouse build by GRO and shows an overview of activities at each. This includes the current crop, expected harvest date, and most importantly, the date of the last visit from GRO staff along with the date of the next planned visit. While it does replicate information found on both the paper worksheets and online data management system, the board serves as a much easier to access visual

snapshot of current operations. As GRO expands and builds more greenhouses, this board has become a valuable planning tool. The WhatsApp group has also improved communication within the team while also aiding in reporting from field staff during their day-to-day activities. This group was created after observing almost ubiquitous use of WhatsApp among the GRO staff and was quickly adopted. The group is now used daily to coordinate activities across different regions of the country and share quick reports and pictures from the field. The introduction of these two informational tools has greatly improved communication and planning among the GRO field staff and is highly recommended to other organizations operating in similar contexts.

Overall the design and implementation of a lean data management system by GRO Greenhouses can be considered a success. The use of lean data principles was especially important for GRO as a robust, efficient system needed to be put into place quickly. By analyzing current GRO processes, the data that needed to be collected at each stage of a greenhouse's lifecycle were identified. This process mapping also allowed GRO to evaluate and improve operational efficiency and is highly recommended for other organizations interested in creating their own data management systems. By identifying the specific data that were essential for operations and removing all other questions, GRO has also been able to ensure that interactions with farmers stay as natural as possible without injecting obvious data collection. Data collected for the GRO lean data system are already a part of the conversations that take place between GRO staff and farmers. Applying the lean philosophy to data collection has resulted in a system that collects and aggregates information from these conversations in a systematic fashion without requiring farmers to answer repetitive survey questions.

However, care should be taken to not change existing workflows too much, as this can result in the rejection of the new system by current staff. This was one of the motivations behind creating a system that utilizes both paper and digital forms— recording data on paper was already a common part of the workflow. Finally, a continuous dialogue with field staff is necessary when creating a system like this. Field staff members are the individuals who are ultimately responsible for data collection and without their acceptance, a data collection system is more likely to fail. Keeping an open dialogue with these individuals allow for the incorporation of their feedback into the system thus encouraging ownership while also gathering the unique invaluable insights that field staff will have. This dialogue does not stop after the first iteration of system design and development, but rather is ongoing throughout the life of the data management system.

VIII. Conclusion

This article presents a case study of GRO Greenhouse and the lean data management system that was created for GRO's operations. The system was built using Google Documents, an off-the-shelf platform that allowed for the creation of robust, cloud-based data collection and analysis tools. This system is coupled with paper worksheets that can be used in the field and then later digitized into GRO's system. Lessons learned during the creation of this system were discussed and it is hoped that other organizations can use this article as a guide for the creation of their own custom lead data management solutions. The creators of humanitarian data collection tools like ODK and KoBoToolbox might also use some of the insights presented here when creating future versions of their tools.

IX. Works Cited

[1] Martin Hilbert, "Big Data for Development: From Information- to Knowledge Societies," 2013.

[2] Linnet Taylor and Ralph Schroeder, "Is bigger better? The emergence of big data as a tool for international development policy," *GeoJournal*, vol. 80, no. 4, pp. 503-518, 2015.

[3] Matti Keltanen. (2013, April) Why 'lean data' beats big data. [Online]. http://www.theguardian.com/media-network/media-network-blog/2013/apr/16/big-data-lean-strategy-business

[4] Kendra Leith and Elizabeth Hoffecker. (2016, February) NexThought Monday – Lean Research: Introducing a Movement for Change. [Online]. http://nextbillion.net/nexthought-monday-lean-research-introducing-a-movement-for-change/

[5] Sasha Dichter, Tom Adams, and Alnoor Ebrahim, "The Power of Lean Data," *Stanford Social Innovation Review*, Winter 2016.

[6] Carl Hartung et al., "Open Data Kit: Tools to Build Information Services for Developing Regions," in *ICTD 2010*, London, 2010.

[7] Rohit Chaudhri et al., "Open data kit sensors: mobile data collection with wired and wireless sensors," in *Proceedings of the 2nd ACM Symposium on Computing for Development*, Atlanta, 2012.

[8] Nicola Dell, Nathan Breit, Timóteo Chaluco, Jessica Crawford, and Gaetano Borriello, "Digitizing Paper Forms with Mobile Imaging Technologies," in *Proceedings of the 2nd ACM Symposium on Computing for Development*, Atlanta, 2012.

[9] Kuang Chen, Akshay Kannan, Yoriyasu Yano, Joseph M. Hellerstein, and Tapan S. Parikh, "Shreddr: pipelined paper digitization for low-resource organizations," in *Proceedings of the 2nd ACM Symposium on Computing for Development*, Atlanta, 2012.

[10] Google. Query Language Reference (Version 0.7). [Online]. https://developers.google.com/chart/interactive/docs/querylanguage