

**WEFTA: Data Organization and Visualization for the Water Engineers for the Americas and Africa**



by

**Andrew Brush**

**Kristina Derrivan**

**Michael Doucette**

**Meghan Lyle**

**WEFTA: Data Organization and Visualization for the Water Engineers for the Americas and Africa**

An Interactive Qualifying Project  
submitted to the Faculty of  
WORCESTER POLYTECHNIC INSTITUTE  
in partial fulfillment of the requirements for the  
degree of Bachelor of Science

by  
Andrew Brush  
Kristina Derrivan  
Michael Doucette  
Meghan Lyle

Date:  
13 December 2023

Report Submitted to:

John DiRuggiero  
Water Engineers for the Americas and Africa

Gar Clarke  
GeoConSults

Professors Golding and Solomon  
Worcester Polytechnic Institute

This report represents the work of WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on its website without editorial or peer review. For more information about the projects program at WPI, please see <http://www.wpi.edu/academics/ugradstudies/project-learning.html>

## **Abstract**

Our project goal was to develop a methodology to improve the collection and storage of data regarding current and future Water Engineers for the Americas and Africa (WEFTA) projects which is currently collected and stored inconsistently. We created multiple dashboards, a database model, and a data entry form. We developed dashboards in ArcGIS and Tableau respectively. WEFTA is currently vested in ArcGIS, although they may consider exploring Tableau and a database management language, such as SQL, as more suitable options. Regardless of the software used, we recommend WEFTA display ten KPIs in the dashboard, including project location, costs, descriptions, and coordinators, as well as community population and number of healthcare facilities.

## **Acknowledgements**

Special Thanks to our Sponsors John DiRuggiero and Gar Clarke,  
WEFTA staff Carla Connors and Paul Kennedy,  
and GIS Coordinator Larry Brotman

## Executive Summary

Water Engineers for the Americas and Africa (WEFTA) is a non-profit organization headquartered in Santa Fe, New Mexico that helps provide needed water and sanitation resources to communities in the Americas and Africa. They have designed and completed 120 projects across the globe and many more are currently underway (WEFTA n.d.). WEFTA continually monitors project sites, but data collection among sites differs significantly. This data is crucial in determining the status of an individual project as well as monitoring the organization as a whole.

Our goal was to create a model methodology to standardize data collection and storage, as well as create a dashboard to visualize this collected project data more efficiently. We followed multiple objectives to achieve this goal:

1. Determined best practices for data entry, storage, and display;
2. Determined WEFTA needs and preferences;
3. Developed a prototype dashboard, database, and data entry form;
4. Considered feedback from the sponsor and other professionals; and,
5. Redesigned based on the feedback from the previous objective.

We surveyed WEFTA volunteers, staff members, and board members using Survey123 to gauge what data points and KPIs are important relative to their position at WEFTA. We created multiple dashboards to visualize different types of project data and make them more actionable for all levels of WEFTA staff. We also created a model database, a data dictionary, and data entry form that can be further populated with relevant data and customized as WEFTA needs grow.

As we began to explore current WEFTA needs and practices, we concluded that the current WEFTA methodology had room for optimization. Due to the nature of WEFTA projects, data collected in the country is often sporadic and not standardized. Certain metrics are specific to certain projects, and this makes organizing, visualizing, and tracking the status of a specific project site difficult and often disorganized. Since there is a surplus of project data, the organization does not intend to organize any past data, which primarily exists as narrative trip reports. Instead, WEFTA would like to organize any future project data moving forward and have a model to develop as their organization expands.

### Recommendations for Data Collection

**1.1** Since there is no standard for data entry, we suggest implementing a more optimized system of data collection in a data entry form, as well as a place to organize, store, and access the data. WEFTA currently uses ESRI products as its focal software, primarily Survey123 and ArcGIS. WEFTA's current methodology is to store survey responses in a spreadsheet and table in an ArcGIS dashboard. After researching ESRI's capabilities and

speaking with experts, we have concluded this is less than optimal and that other software for collecting and storing data may be more beneficial to the organization's functioning and tailored to their current needs.

## **Recommendations for Data Storage**

**2.1** Our recommendation is to store this data in a database created by a database management programming language so that the most control can be exhibited over the structure of the storage. We also recommend using a different method of data collection instead of storing data in a spreadsheet, as Survey123 would be inefficient when working with our database design. Our proposed database was developed in SQL but query languages like Malloy, Basis, or broader languages like Python or Java would achieve the same goals that we aimed to model in SQL.

**2.2** While developing the database, we decided to structure the project key to capture enough information to identify the project and location but to avoid overuse of embedded information. We believe that this key is the best option given the scope of our project. However, it is our recommendation that WEFTA invest in further research into encryption, data security, and algorithmic value generation and decryption to streamline the process by which data is inserted into the database.

**2.3** We have concluded based on discussion with our sponsor that the best way of sustaining our database design and making it accessible to people new to working with it is to make a data dictionary. We developed that deliverable in line with sponsor examples.

## **Recommendations for Data Visualization**

**3.1** The use of KPIs would greatly increase WEFTA's ability to condense metrics into quantifiable data that can be used to monitor the status of projects across the globe accurately. Based on our research, we determined that no more than ten KPIs can be displayed on each dashboard, as the human short-term memory can process  $7 \pm 3$  components at once. Because of this, we recommend using the top ten most important KPIs, as determined from our needs assessment survey, in the general dashboard. These KPIs are project description, photos, total population, number of healthcare facilities, project cost, coordinator contact information, partner/community contact information, project type, community engagement, and total donations. This will ensure the most pertinent information is accessible for all members of WEFTA.

**3.2** We also recommend using all of the available financial KPIs within the financial dashboard. There are only seven financial KPIs currently, so they all can be displayed based upon the findings mentioned previously.

**3.3** Currently, WEFTA prefers to collect, hold, and analyze its data using ArcGIS and associated ESRI software, but the organization uses only some of ArcGIS's capabilities. We recommend using a software package that is less geospatially tailored and more focused on data visualization for WEFTA dashboards. Due to Tableau's enhanced visualization functionality and status as an industry standard, we recommend WEFTA consider using the software for the development of future Dashboards. We researched alternative dashboard software and recommended WEFTA continue to assess alternative software should Tableau prove inadequate or inappropriate. Due to WEFTA's familiarity with ESRI products, we recommend further research into data transfer and compatibility between ESRI Arc products, Tableau, and SQL as well.

## Authorship

Section	Author(s)	Editor(s)
Abstract	MD	ML, MD
Executive Summary	KD	ML, MD
Introduction	MD	AB, KD, MD
<b>2.0 Background</b>		
2.1	MD	AB, MD
2.2	ML	ML
2.3	AB	AB
2.4	KD	AB, KD
<b>3.0 Methods</b>		
3.1	KD	AB
3.2	MD	AB
3.3	KD, ML	AB, MD
<b>4.0 Findings</b>		
4.1	MD, ML	KD, MD, ML
4.2	AB, ML	AB, KD
4.3	ML	KD
4.4	MD	KD, MD
4.5	AB, KD	KD, AB
4.6	AB, KD	AB
<b>5.0 Conclusions</b>	<b>AB, KD, MD, ML</b>	<b>AB, KD, MD, ML</b>
Appendix A	MD	MD
Appendix B	MD	MD
Appendix C	KD, AB	AB, KD
Appendix D	KD, AB	AB, KD



## Table of Contents

Abstract.....	i
Acknowledgements.....	ii
Executive Summary .....	iii
Authorship.....	vi
Table of Contents.....	vii
List of Figures .....	viii
List of Tables .....	ix
1.0 Introduction .....	1
2.0 Background.....	3
2.1 WEFTA’s Mission and Approach .....	3
2.2 Data Organization.....	4
2.3 GIS Data .....	7
2.4 Database Accessibility .....	9
2.5 Database Sustainability .....	10
3.0 Methods .....	13
3.1 Best Practices.....	13
3.2 Determine Needs and Preferences.....	13
3.3 Prototype Development and Testing.....	14
4.0 Development of the Deliverables and Findings .....	15
4.1 Data Organization and Needs .....	15
4.2 Database.....	15
4.3 Data Storage .....	16
4.4 Needs Assessment Survey .....	23
4.5 Dashboard .....	24
4.6 Documentation and User Training .....	30
5.0 Conclusions and Recommendations .....	32
6.0 References.....	34
Appendix A: Needs Assessment Survey .....	36
Appendix B: Contact Information Survey.....	37
Appendix C: Interview Questions .....	38
Appendix D: Dashboard Training Manual.....	40

## List of Figures

Figure 1.01 The proportion of safely managed drinking water services by country in the world ..1	
Figure 2.01 The flow of a WEFTA project .....	4
Figure 2.02 Examples of various GIS layers .....	7
Figure 2.03 Project Images from Honduras within Storymaps on the WEFTA website .....	9
Figure 3.01 Spreadsheet depicting how some WEFTA data is currently stored (fake data) .....	14
Figure 4.01 Diagram of proposed database schema .....	28
Figure 4.02 Benefit of multi-tabular design .....	30
Figure 4.03 The portion of the data entry form in which project data is entered .....	31
Figure 4.04 Color coded diagram demonstrating ProjectID generation .....	32
Figure 4.05 Project table of data dictionary .....	33
Figure 4.06 Chart of KPIs with highest amount of “Very Important” Responses .....	35
Figure 4.07 Example of Tableau Viz-in-Viz feature .....	37
Figure 4.08 Final Tableau Overview Dashboard.....	38
Figure 4.09 Final Tableau Financial Dashboard .....	39
Figure 4.10 Final ArcGIS Overview Dashboard .....	40
Figure 4.11 Final ArcGIS Financial Dashboard .....	41
Figure 4.12 An excerpt from dashboard documentation .....	42

## **List of Tables**

Table 1.01 A theoretical database relation.....	6
---	---

## 1.0 Introduction

The lack of improved water, sanitation, and hygiene resources (WASH) in developing countries has had significant health impacts on one-third of the world's population. More than two billion people lack access to safe drinking water, and another 675 million people still practice open defecation which leads to water contamination and is responsible for the deaths of 1,000 children a day in regions with low WASH resources (UNICEF 2021).

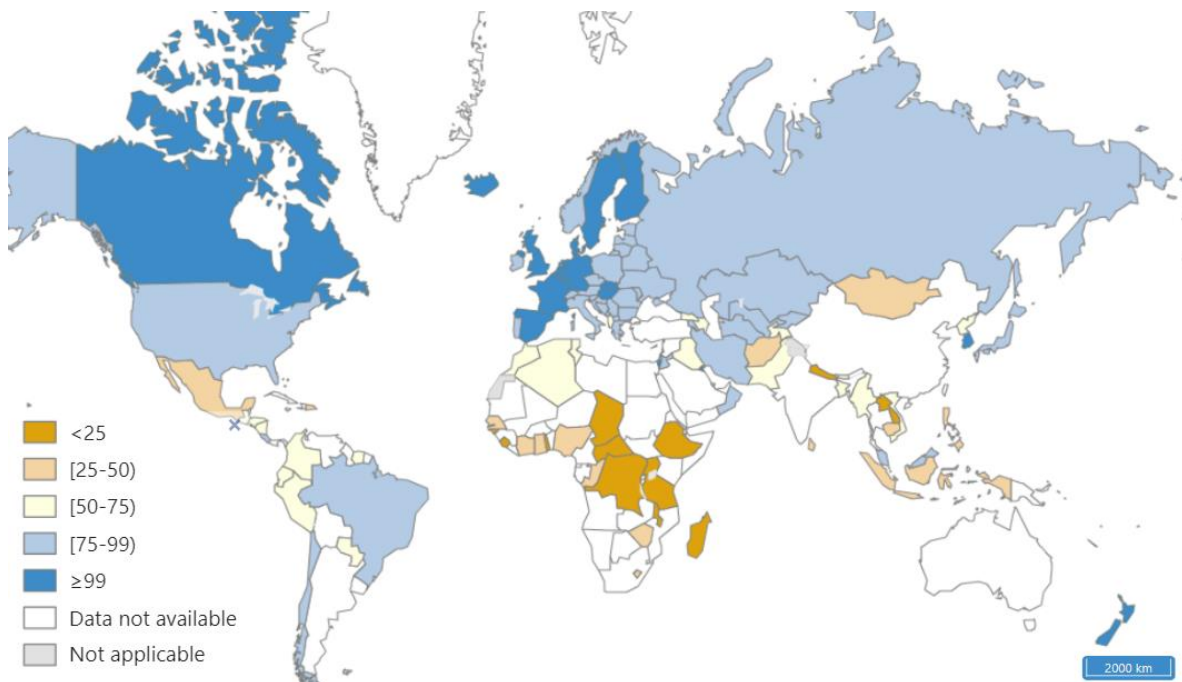


Figure 1.01 The proportion of safely managed drinking water services by country in the world

WEFTA is a non-profit organization headquartered in Santa Fe, New Mexico that helps bring WASH resources to underprivileged communities. WEFTA strives to bring these basic resources to underdeveloped regions because WASH resources are a basic human need that have significant health benefits (Tseole et al., 2022). Proper access to WASH resources can lead to a ten percent reduction of the global disease burden (UN n.d.). WEFTA has set up numerous projects across Latin America and Africa, and has “successfully completed more than 120 projects, positively impacting over 195,000 lives” (WEFTA n.d.).

The goal of our project was to develop a methodology to improve the collection and storage of data on current and future WEFTA projects so that they can be more accessible for WEFTA board, staff, and volunteers. We achieved this goal by meeting the following project objectives:

1. Determined best practices for data entry, storage, and display;
2. Determined WEFTA needs and preferences;
3. Developed a prototype dashboard, database, and data entry form;
4. Considered feedback from sponsor and other professionals; and,
5. Redesigned based on the feedback from the previous objective.

We surveyed WEFTA volunteers, staff members, and board members using a Survey123 form to gauge what data points and KPIs are important to them and what is important for them to see in a dashboard relative to their position at WEFTA. We created multiple dashboards so that WEFTA volunteers, staff, and board members can monitor the status of projects. We also created a database model, a data dictionary, and a prototype data entry form.

## 2.0 Background

In supporting WEFTA's data storage structure and methods for visualizing the data stored, we first researched WEFTA's current goals, the theory behind and benefits of efficient data organization, the systems administrative software that WEFTA is the most familiar with and invested in (ArcGIS and other ESRI products), the industry standards in creating an accessible database, and common practices for promoting database sustainability.

### 2.1 WEFTA's Mission and Approach

WEFTA's mission is primarily to "connect water professionals and partners with communities in developing countries to improve water, sanitation, and hygiene resources for the benefit of the people and the environment" (WEFTA n.d.). WEFTA is funded through donations and grants where up to eighty percent of those funds are allocated to WASH projects in Latin America and Africa. Notable examples of some WEFTA projects are overseeing the installation of water pipes in Honduras, developing an electric pump in Colombia, and providing chlorine water treatment and guidance on maintenance for a healthcare facility in Ethiopia. The Honduras project was created when community water operators reported lack of flow from their water distribution system. A WEFTA volunteer went to investigate the problem, which was determined to be from water thieves taking water from the system and leaving air valves open. The volunteer then conducted training on how to use the system properly at three local villages. Furthermore, the volunteer, with the help from some locals, laid and updated more piping around the area to increase the range and distribution of clean water (WEFTA n.d.).

A WEFTA project typically follows the sequence illustrated in Figure 2.01. First, a request is submitted by a community or another organization, which is then reviewed by WEFTA's board of directors<sup>1</sup>. If it is approved by the board, a team of volunteers and specialists<sup>2</sup> is sent to the project location (WEFTA n.d.). When the team arrives on the site, they face several problems they hope to overcome such as "geographical inequalities, climate change, investment in WASH resources, low levels of knowledge on water diseases and ineffective local community

---

<sup>1</sup> WEFTA's board consists of seven members including a president, vice president, secretary, and treasurer. They also sometimes act as normal staff members with some personally visiting some of the project sites.

<sup>2</sup> The volunteers come from WEFTA partners and have engineering experience and skills related to the project they are working on.

engagement” (Tseole 2022). They start by conducting a land review to find solutions to the area’s WASH problems. The team proposes a plan and design solutions and the project begins construction. The team oversees and manages the installation and implementation of the project. Once the project is complete and the team leaves the site, WEFTA keeps up to date with the maintenance of the project and its operation.

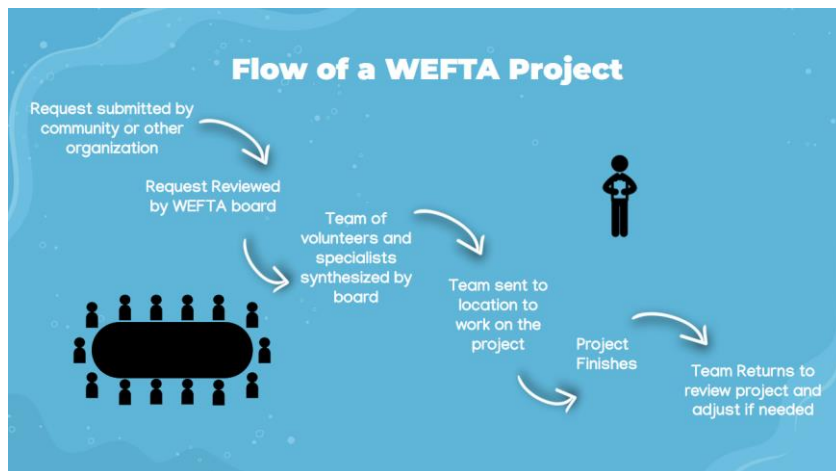


Figure 2.01 Diagram showing the flow of a WEFTA project

## 2.2 Data Organization

According to System Administrator, John DiRuggiero (personal communication, 8 September 2023) disorganized project data is one of WEFTA’s major weaknesses. Most of the data about their projects is loosely defined and exists mainly as diary-like reports and in human memory. This arrangement makes comparison between projects nearly impossible and risks the loss of institutional knowledge when a staff member or volunteer leaves. Improved abilities to compare project data will allow WEFTA to improve efficiency by learning from past experiences. Additionally, by compiling, organizing, and displaying data in condensed statistics, WEFTA can more clearly advertise themselves to potential funders.

Project-based organizations such as WEFTA have the unique issue of project customization—where a project is fitted to address different situations as they arise—which seemingly makes it difficult “to learn and build up their knowledge capabilities from one project to another” (Meyerson, Weick, & Kramer, 1996, Referenced by Fong Boh, 2006)”. Ignoring relations is an issue when it comes to cross-project references and sharing information with people unfamiliar with WEFTA’s existing data, such as the partners whom WEFTA frequently

collaborates with (Fong Boh, 2006). The idea that the projects are completely divorced from one another is a “well-recognized” misconception since “common problems may be encountered across different projects, and effective sharing of knowledge across projects can reduce the organizational costs of duplicating efforts to invent the same solutions (Goodman & Darr, 1998, Referenced by Fong Boh, 2006).” Optimizing finances is a particular goal for WEFTA according to Gar Clarke, New Mexico Geospatial Information Officer (personal communication, September 11, 2023). Optimization calculation becomes far easier when the data is numerically encoded in one computer system in a consistent way, allowing the computer to plug in the correct values into predetermined functions.

Moreover, Diana Chien and David Larson of the MIT Communications lab explain that data housed in disorganized files instead of a database or organized system can be a hindrance to efficiency (Chien & Larson, n.d). Extracting data from WEFTA’s file archive and organizing them into a database table will cut down on the time it takes to find it. Because of the way that the data currently exists, this is an issue as someone wishing to compare projects must parse entries in their entirety. It takes high-level computation to calculate and search data from a report, but preprocessed numbers are much easier for a computer to plug into predictive analysis because no distinction has to take place. It is more useful if the data is broken down into pieces and stored in a predictable and organized system rather than unprocessed in a narrative. Surveys are a useful way to centralize this inconsistently stored data. Each survey submission represents a WEFTA project in terms of the qualities we deem significant.

In addition to compiling all the data in one place, surveys condense data to only its most important information while simultaneously maintaining an acceptable format so that it can be entered into formulas without preprocessing. For instance, each date in Table 1.01 is formatted consistently so that a computer can easily identify the month if we predetermine the month as the first number with no corresponding day. Many survey creation platforms such as Survey123 also have a feature that inputs each submission into a database table where each row, also known as a tuple, represents one WEFTA project and its attributes. For example, Table 1.01 shows how one row indicates a project and each columns show the data that corresponds with that project where all the information in that row corresponds to the same category, either the date, country, latitude of water source, or longitude of water source for each project. In that regard, a database can store



geospatial information. Each row may be representative of one survey submission. One column represents one attribute over each project. Each row is identifiable by a unique code called a “primary key”. Primary keys make it possible to very quickly query, organize, and analyze projects just by calling their names (Ullman & Widom, 2008 p. 23-27).

Primary keys also support Fong Boh’s claim that the relationship between projects and tables exists and is valuable. When a primary key is referenced in the attributes of another database table, this is called a foreign key. Foreign keys “represent the paths along which related tables are joined together in the database (IBM, 2021)” and can be used to show relationships, hierarchy, and dimensionality by nesting database tables through the primary keys of its data. A necessary task would be to decide on a process that determines what the key will be. One option is to take inspiration from the UN’s humanitarian project naming convention that identifies project as a combination of abbreviated representations of its attributes (John DiRuggiero, System Administrator, 8 September 2023). Table 1.01<sup>3</sup> identifies *project name* by the first two letters of the country it takes place in and a number indicating the year the project started. In this case, *project name* is the primary key and is thus underlined. This conveniently compacts information and makes it easier for outside organizations to integrate into the system as it is already familiar in the UN system. Standardizing the data in a familiar and consistent way will improve communication with sponsors and donors about project progress as well as allow WEFTA to gain financial insight quickly and accurately through computational methods.

<u>project name</u>	date	country	latitude	longitude
KE2023	3/1/2023	Kenya	-1.35	36
PA2004	9/20/2004	Panama	8.98	-79.5
KE2012	4/17/2012	Kenya	-1.03	37.05
ME2012	12/18/2012	Mexico	19.4	-99.2

Table 1.01 A theoretical database relation

---

<sup>3</sup> All data in Table 1.01 is hypothetical.

### 2.3 GIS Data

Because each WEFTA project has a distinct geographic location, the data will be integrated into a geographic information system (GIS) so that WEFTA is able to easily record, retrieve, and map data; this will also allow for optimal visualization of the data due to its geographic component. A GIS is a computer system that is primarily used to analyze and visualize geographical data. The data that is utilized within a GIS is made up of two components; a geographical or spatial component and an attributes component which is the data referenced by the geographical component (Maguire, 1991). A GIS visualizes this data through layers. Layers are visual representations of geographical data (e.g., roads, topography) that are placed on top of one another to create a comprehensive and informative model. For example, a basic map may be used as a base layer of topography for visualization purposes. Then, layers are added containing relevant data such as landmarks, water features, and elevation. All of these layers are combined to create an informative geographical visualization of the provided data. As illustrated in Figure 2.02, each layer has specific geographic information, and no single layer provides a full representation of all the data. However, once all the layers are put together, they create a comprehensive image of the geographical data they each store.

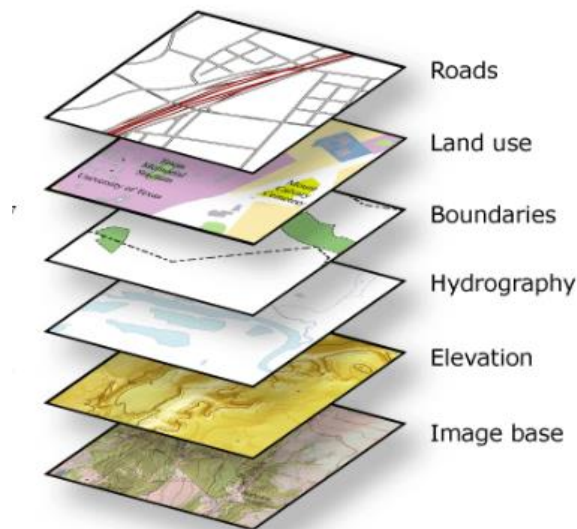


Figure 2.02 Examples of various GIS layers (ArcMap)

Due to their familiarity with the software, their existing paid license, and the software's features, WEFTA would prefer to use ArcGIS's capabilities for their project databases and dashboards. WEFTA already utilizes StoryMaps which is integrated with ArcGIS. StoryMaps is

a web-based map tool that displays a map along with contextual images and videos for certain locations. As shown in Figure 2.03, StoryMaps can display what working on a project with WEFTA looks like. It can also display data visualizations and other information about the projects which can be helpful for outreach and fundraising. One unique advantage about ArcGIS is it contains database and database management system (DBMS) capabilities. A database management system is a system or software that provides functionalities such as data retrieval and manipulation. Databases on their own are simply tables or groups of data; database management systems allow for user interaction. While its database capabilities are more limited than those of dedicated database software, ArcGIS's database provides other benefits such as easy integration with their other applications like Survey123. Survey123 is a tool integrated with ArcGIS that allows for the creation of standardized data entry surveys that link directly to the database. This will allow us to create data collection surveys that seamlessly feed new and standardized data into the database. Additionally, ArcGIS provides information to smooth the process of integrating maps and other data visualizations into websites. Also, having the database within ArcGIS will help streamline and organize the project as a whole. Another benefit of using GIS map visualizations is that they are easy to read and appeal to a wide audience. While graphs and other types of data visualization are very informative as well, they can often be intimidating for those who are not familiar with them. On the other hand, most people are comfortable using and reading maps.

One goal of this project is to create a map with all project locations pinned on the map and project data available when the project pin is clicked on. This way, the data is clearly organized and displayed by project.

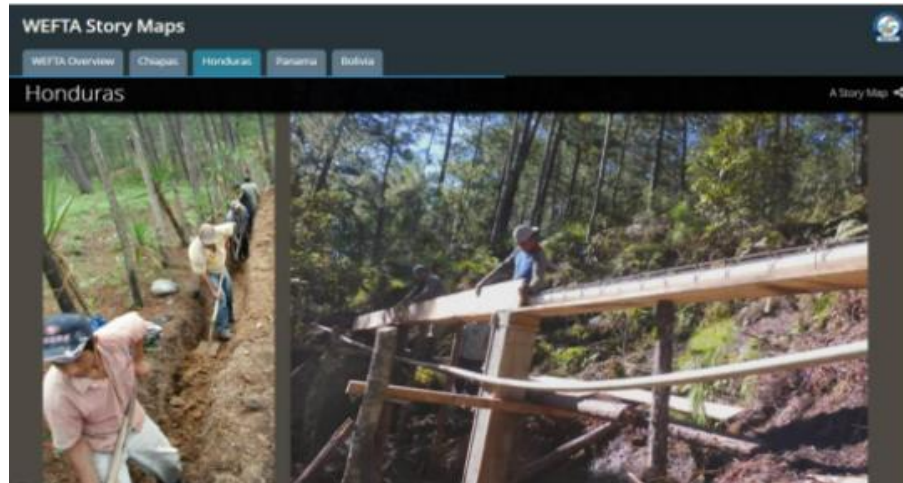


Figure 2.03 Project images from Honduras within StoryMaps on the WEFTA website (WEFTA, n.d)

## 2.4 Database Accessibility

A front-end dashboard for a database can serve as a progress overview for each project depending on the user level: Board Member, Volunteer, or Public Viewer. Front-end development is critical as it acts as the first point of contact for information, facilitates consumer relations, and efficiently organizes relevant data and action items. WEFTA currently has a pseudo-dashboard in the form of an interactive Excel spreadsheet, however, the layout is very crowded and not integrated into ArcGIS.

An article from the journal *NonProfit Quarterly* discusses the main role of a nonprofit dashboard, which is to organize data and key performance indicators (KPIs) in a visually appealing manner and highlight the relationship between those key indicators. A dashboard also serves a secondary role as a snapshot of the current status of the organization in order to allow for a periodic organizational progress assessment, as a way to highlight any improvements or action items (Suarez et al., 2020).

Based on the current dashboard, the identified KPIs include a site performance scoring system, with each project center's irrigation system components being rated on a Likert scale of 1-5, which then contributes to an overall score out of 100. Qualitative scoring systems are shown to be vague and often arbitrary (Bishop, 2015), and the Likert system currently instituted is not the optimal way to present the status of each project site component, which is a key goal of a non-profit dashboard. Verdinelli & Scagnoli (2013) show that in a sample of three qualitative research journals, the most common industry method for displaying this type of qualitative data

was a matrix display, being used 60.2% of the time to display and analyze data. A matrix display shows the frequency of certain phrases and response patterns compared to the whole.

This method will be most beneficial for a more defined organization of status descriptions and can be used in conjunction with the already imposed 1-5 Likert scale method, to preserve the ‘big-picture’ grading system currently instituted, while still allowing streamlined access to what those grades mean qualitatively. Combined, these two visualization methods will encompass both quantitative and qualitative data and span the entirety of the data collected in a typical WEFTA survey.

## 2.5 Database Sustainability

To ensure the database continues to be useful in the future, it must be created with sustainability in mind. To ensure the database's sustainability, standardized methods need to be created for use moving forwards. The database must be monitored based on a “commonly accepted and well-documented methodology” that is standardized such that there is little room for interpretation (Granke, 2013). Due to the nature of the data that will be collected, certain aspects will likely have to contain some level of interpretation. However, this must generally be kept to a minimum. This will ensure the consistency and quality of data entering the database. The methodology for data monitoring and insertion is vital for the sustainability of the database because if different members of staff are able to interpret it differently, inconsistent data may begin to fill the database. This would ultimately lead to the database becoming potentially useless for WEFTA. Also, to further ensure data quality, conformity checks can be put into place. These checks range from single and multi-parameter checks to spatial comparisons. Conformity checks are automated and test for errors to “validate the plausibility of data” (Granke, 2013).

Employee training has been shown to be critical in organizational performance and the overall success of a company. When employees feel competent and supported in their responsibilities, not only is the organization able to function more efficiently, but employee morale is also shown to improve (Walters & Rodriguez, 2017). As part of the long-term upkeep of the database, WEFTA employees must be trained on how to use the database so information can be uploaded and formatted into the database properly. If the data is not formatted properly it can lead to issues when displayed and overall impede the communication of data to stakeholders

and potential beneficiaries. Additionally, employees must be equipped to diagnose and fix any future issues, as well as have the ability to facilitate database expansion if needed. Since the database will be mostly in-house, our training material is constrained to how to use the ArcGIS software, as well as its compatible apps, like Survey123. ArcGIS has training videos on the general use of ArcGIS, but a significant portion of ArcGIS's capabilities will go unused, or only in rare and specific instances. Due to this, utilizing the pre-made ArcGIS training videos will be inefficient and use up the limited staff training time when in Santa Fe. A more specialized set of training modules will be a better use of the time constraints given.

There are many formats of training, and their effectiveness is heavily dependent on the context of the training material and the constraints of the delivery, as shown in Martin et al. (2013). According to a review of all training methods, the methods most relevant to ArcGIS training will be stimulus-based training and Programmed Instruction-based training. Programmed-Instruction-based training "Involves the delivery of training through instruction that is delivered by a program via some electronic device *without* the presence of an instructor; the electronic device can be a computer, DVD player, CD player, etc." (Martin et al., 2013). Programmed instruction allows for a more flexible training schedule, as well as giving the trainees the opportunity to re-watch the training if they encounter issues in the future. The online nature of the Programmed-Instruction-based training also allows for the ability for more visually varied and interactive content, which is shown to increase overall retention. This method does require employees to take the initiative to pay attention and learn the database, which is often considered a dry subject and may be hindered by the decision to cheat or skim through training modules.

Another training method, Stimulus-based training, is defined as "Using some type of stimulus (i.e., music, works of art, narratives, etc.) to motivate the learner to learn. The training induces a state of being (e.g., relaxation or awareness) in the participants to achieve learning" (Martin et al., 2013). It is often combined with stressful or 'dry' material to facilitate a positive and more effective learning experience. The example given in Martin et al. (2013) is utilizing music to lessen human failure cues when learning mathematics. The downside to this method is that its unconventional nature is a deterrent for some users, and they may take training less seriously.

There is not an all-encompassing training method for database upkeep, and each method has significant benefits and disadvantages. A combination of multiple methods will allow for a realistic dissemination of information within the WEFTA time and financial constraints, while also allowing for the benefits of increased employee retention and competency.

ArcGIS has numerous tools tailored to data collection and visualization that suit WEFTA's organizational needs, such as Survey123 integration, StoryMaps, and individual data layers. For standardization, the UN naming convention is shown to be the industry standard and also covers WEFTA's overall needs. Data visualization is also shown to be critical for an organization's daily function and can be streamlined with a dashboard and user-tailored modules. Industry standards will be the foundation of both the dashboard and database, but the specific implementation of these industry standards for WEFTA systems may require adjustment due to WEFTA-specific needs and will be addressed further.

### 3.0 Methods

The goal of our project was to develop a methodology to improve the data collection and storage of current and future WEFTA projects so that they can be more accessible for the WEFTA board, staff, and volunteers to view. Our framework was as follows:

1. Determined best practices for data entry, storage, and display;
2. Determined WEFTA needs and preferences;
3. Developed a prototype dashboard, database, and data entry form;
4. Prototyped deliverable testing; and,
5. Redesigned based on the feedback from the previous objective.

#### 3.1 Best Practices

We conducted research into how WEFTA currently collects, stores, and displays data, and researched how other companies and organizations handle data similar to what WEFTA is currently dealing with. Our initial research showed us that continuing the use of ESRI and the software that they offer would be in the best interest for WEFTA due to the organization's expressed familiarity with the software; however, we discovered that the software Tableau was the industry standard for dashboards and decided to develop dashboards in both ArcGIS and Tableau to explore different capabilities.

#### 3.2 Determine Needs and Preferences

Our sponsor had an expressed interest in having multiple dashboards that display different KPIs for the different members of the WEFTA organization. To determine how to group the KPIs for each dashboard and how many unique dashboards we would create, we sent out a Survey123 form (see appendix A) to the WEFTA administration to distribute amongst WEFTA personnel. This form divided each data point into four sets: general project information, community information, finance, and technical information. A matrix was also included where the survey taker could rank the importance of each data point to their position by choosing from the options of most important, somewhat important, least important, or an indication of not understanding the datapoint. We also asked for information about the user's position in the organization as well as their experience and contribution in order to gauge which responsibilities correspond with which data points. This survey also included a link to a separate survey that



collected contacted information from the respondents separate from their responses for future testing on the prototype dashboards.

### 3.3 Prototype Development and Testing

We developed, tested, and revised three major deliverables: a database model, a data entry form, and a set of dashboards. Figure 3.01 depicts an example of how some data is currently stored at WEFTA and part of the basis for our deliverables.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Country	Community	Project_ID	Project_Category	Completion_Date	Description	Total_Cost_USD	Project_Status	Local_Partner	Served_Population	longitude	latitude	Healthcare_Facility_Scores
2													
3	Bolivia	Kerani	123_32_56	Water	11/1/2004 0:00	Gravity fed water system	24265	Completed	Suma Jayma	959	-68.51	-16.1555	5
4													
5	Bolivia	Hichuraya	123_32_57	Water	11/1/2005 0:00	Gravity fed water system	13559	Completed	Suma Jayma	385	-68.0494	-16.9346	3
6													
7	Bolivia	Villa El Carm	123_32_58	Water	12/1/2006 0:00	Gravity fed water system	12155	Completed	Suma Jayma	245	-68.5563	-16.9416	4
8													
9	Bolivia	Laca Laca	123_32_59	Water	6/1/2007 0:00	Gravity fed water system	11120	Completed	Suma Jayma	189	-68.1161	-16.8867	1
10													
11	Bolivia	Huacallaya	123_32_60	Water	12/1/2007 0:00	Gravity fed water system	10654	Completed	Suma Jayma	175	-68.4495	-16.9189	2
12													
13	Bolivia	Playa Verde	123_32_61	Water	2/1/2008 0:00	Individual hand pumps project	5351	Completed	Suma Jayma	245	-68.5464	-16.3006	5
14													
15	Bolivia	San Pedro	123_32_62	Water	2/1/2008 0:00	Individual hand pumps project	4395	Completed	Suma Jayma	203	-68.5538	-16.294	3
16													
17	Bolivia	Waillani	123_32_63	Water	2/1/2008 0:00	Individual hand pumps project	4684	Completed	Suma Jayma	217	-68.5623	-16.2962	4
18													
19	Bolivia	Pata Paya	123_32_64	Water	4/19/2008 0:00	Individual hand pumps project	6048	Completed	Suma Jayma	196	-68.5584	-16.3047	1
20													
21	Bolivia	Huallatiri	123_32_65	Water	5/30/2008 0:00	Gravity fed water system	20628	Completed	Suma Jayma	161	-68.5177	-16.9318	2
22													
23	Bolivia	Chacoma Irp	123_32_66	Water	11/30/2008 0:00	Gravity fed water system	31961	Completed	Suma Jayma	651	-68.3123	-16.8256	5
24													

Figure 3.01: Spreadsheet depicting how some WEFTA data is currently stored (fake data)

Once the prototype dashboards were created, we interviewed Project Coordinator Paul Kennedy as he indicated on the KPI survey that he would be willing to be interviewed. We spoke to Mr. Kennedy over a Zoom video call and prompted him through tasks to test the intuitive functionality of both the Tableau dashboard and ArcGIS dashboard. We also asked for his opinion on the dashboard’s aesthetic qualities, notably color scheme and layout. See Appendix C for the questions asked during the interview.

## 4.0 Development of the Deliverables and Findings

### 4.1 Data Organization and Needs

We have found that WEFTA is a very small organization, which made data collection on workplace opinion difficult. Notably, we were only able to contact nine WEFTA members for our KPI survey, of which eight responded. Later, this would prove to impact our dashboard development. It was very difficult to get a broad overview of the general opinions of the organization, as we could only contact a maximum of nine people at a time. Other data not pertaining to opinion was also difficult to gather because of organizational privacy. This led to issues in beginning the deliverable development process. John DiRuggiero provided us with an Excel pivot table of WASH scores, which are values that calculate the level of concern in different areas of WASH for each community served. In addition, we were given a list of data points that John DiRuggiero and Tim Wellman, WEFTA's C.E.O, wished to see stored in a data storage structure. Although this helped us begin to develop a preliminary schema, we had difficulty understanding what certain datapoints were referred to or were restricted to because we initially had minimal examples on which to base our model.

It was also difficult to get test data to display on the dashboard because of privacy concerns. We have learned that certain data is especially proprietary since WEFTA formed a partnership with another organization that prohibited WEFTA from sharing data (personal correspondence, John DiRuggiero, Systems Administrator). All of the data used in the dashboard was fictitious data developed by both John and us. Eventually, we had enough information to develop a dataset. After cleaning and parsing it, we developed the database to capture all possible data points.

### 4.2 Database

We began the development of our database design by using dbdiagram.io, a browser software that takes a coding language similar to SQL as its input and generates a schema diagram.

We grouped data points by combining related ideas into tables and deciding on the linking factors between tables, which would become our primary and foreign keys. We then typed a preliminary version of this model using SQL upon the advice of geospatial professional

Larry Brotman, who emphasized the importance of housing geospatial information exclusively in ArcGIS and all other data in a separate database (personal correspondence, Larry Brotman, GIS Coordinator). Considering advice from correspondent Gar Clarke, we also made a data dictionary (a compilation of all datapoints, their purpose, and format) to clarify our database schema as well as describe each data point to new developers. which was developed as an Excel spreadsheet per an example made by Gar Clarke.

#### 4.3 Data Storage

In light of our conversation with Larry Brotman (New Mexico GIS Coordinator, Government of New Mexico, 31 October 2023), we decided to house data in an SQL database and visualize it in both ArcGIS and Tableau instead of housing it in a spreadsheet connected to a visualization software. We organized the data into interconnected tables instead of the preliminary single table modeled by our sponsor. We made this decision to make the schema clearer to new users as the data is organized by descriptive tables, and the relationships between data points are more apparent. We also decided that dividing the data by categories would make the dataset less dense and easier to traverse. The database diagram (Figure 4.01) was then converted into SQL code to demonstrate the functionality of the diagram.

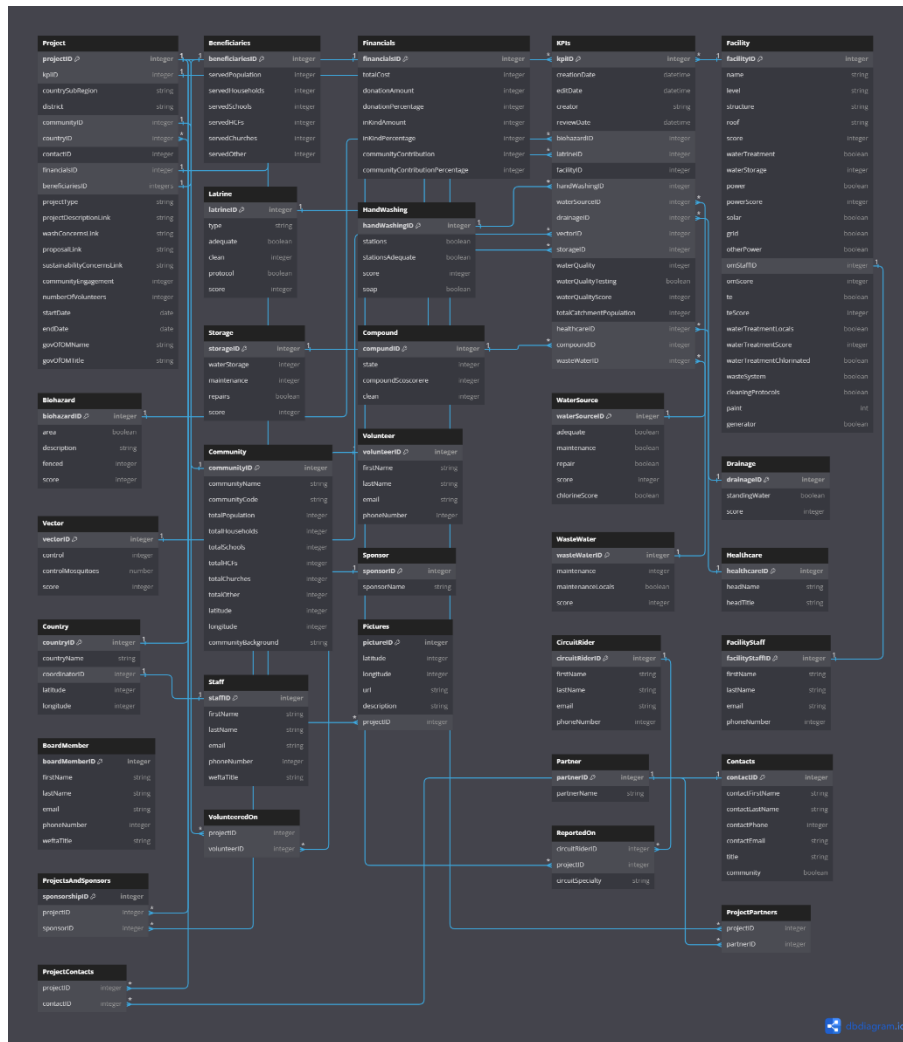


Figure 4.01 Diagram of proposed database schema

WEFTA currently documents projects with free-form written accounts. We considered parsing the accounts and deciding which datapoints were consistently represented overall, but found difficulty in the fact that the accounts were not standard enough to compare them. We then considered ignoring the accounts completely, but decided against that when Project Coordinator, Paul Kennedy, explained that project reports were inspiring to him as a volunteer (personal correspondence, 31 October, 2023). We decided to address this by including a hyperlink to the written report in a field of the corresponding data table.

When deciding what other data to store in the database, we combined two datasets provided by our sponsor. One was the list of data points desired by Tim Wellman, the WEFTA C.E.O, and the other was the Excel pivot table constructed by John. We then chose the distinct

values between the two sets and organized them to make the schema diagram. This new dataset would also be used as the basis of our KPI survey. Initially, we considered using only the calculated fields of the pivot table, which refers to the overall totaled scores of each category but decided to include everything to be more comprehensive.

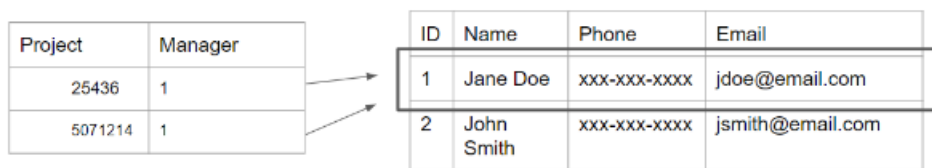
We referred frequently to John DiRuggiero's example of data entry. We used this to determine data types and acceptable values in our data dictionary, as well as to get a sense of what each datapoint in the pivot table was describing. For ease of use in the dashboard, we also had to manually clean the data. Fields that represented a question following the format "if you responded x to the previous question..." were changed to be independent columns. For example "existence of water source" and "if water there is a water source, water source quality" was limited to only "water source quality" as any value other than null for water source quality would imply the existence of the water source. The reason for having the two questions initially was to allow the user to skip questions that do not refer to them. However, our multi-tabled design allows for the user to ignore full tables if they wish as well as accept null values so that it is possible to ignore irrelevant fields.

We also added columns where John DiRuggiero entered multiple values per field. For example, John DiRuggiero's example of "power source" had "solar and grid" as several responses. This would prove to be difficult if we decided to compare facilities that used solar power to facilities that used grid power as grid and solar are included as one answer. To solve this, we added three columns: 'solar,' 'grid', and 'other' that accepted a true or false indicating if the facility used that type of power. We amended the dataset to more closely mirror John's example of data submission as well as our structural changes. We considered normalizing, which is the process of rearranging the schema of the database to reduce anomalies. An anomaly in a dataset is a datapoint that makes the data table too interdependent, and it best practice to resolve because they "often incur increased machine time, loss of business, [and] data inconsistency (Lee, 1995)."

Overall, normalization would have made the database more efficient and closer to industry standard. However, we decided to sacrifice efficiency in order to prioritize maintaining the categorization of the data. While normalization would have been better practice, it would have made the database far less comprehensible. The categorization was useful to us in

contextualizing the data, and we believe that the strict categorization would make the database structure make more logical sense to new users.

We designed the database so that the highest-level project database references other tables with more in-depth information. This makes it easier to ignore values that do not pertain to the project (i.e. latrine information in a water source construction project) and to allow us to reuse values that are used in more than one tuple, such as employee information. This would avoid redundancy and make data entry easier. For example, hypothetical project 25436 served a community in Kenya and was managed by Jane Doe. Our model makes it so that if Jane manages project 5071214 in Panama, she only has to enter her contact information once instead of every time she manages a project. See Figure 4.02 for a visual representation of this concept. We believe that making data entry more convenient will encourage more volunteers to enter data in the database.



This model is more convenient to enter data into, because Jane only needs to type her full name, email, and phone number one time. She also only has to enter two fields at a time per project.

Project	Manager	Phone	Email
25436	Jane Doe	xxx-xxx-xxxx	jdoe@email.com
5071214	Jane Doe	xxx-xxx-xxxx	jdoe@email.com

This model is less convenient because Jane has to type her full name, phone, and email twice.

Figure 4.02 Benefit of multi-tabular design

We also found that the interconnectivity of our database model meant that data would have to be collected in a more systematic and complex way than the Survey123 form we initially proposed. Since this approach would likely be out of the scope of our project, we made a fillable digital document based on the layout of the data dictionary that our sponsor could convert into a user interface of his choice. See Figure 4.03 for a screenshot of the portion of the form in which project data can be entered.

Project		Insert Value
Datapoint	Description	
projectId	Country phone prefix, Community Sequence Number	
kpId	Sequence Number that Identifies KPI table	
countrySubRegion	Name of the sub region	
district	Name of district served	
communityID	Reference to Community Table	
countryID	Reference to Country Table	
contactID	Reference to Contact Table	
financialsID	Reference to Financials Table	
beneficiariesID	Reference to Beneficiaries Table	
projectType	Limited to .....	
projectDescription Link	Link to free form written report on project	
washConcernsLink	Link to document with Water, Sanitation, and Hygiene Concerns	
proposalLink	Link to proposal document	
sustainabilityConcernsLink	Link to document with sustainability concerns	
communityEngagement	Number of community members participating in project	
startDate	Date of touchdown in community	
endDate	Date of takeoff	
govOFOM	Official for operation and maintenance (person name)	
govOFOMTitle	Title of Official for operations and maintenance (person title)	
numberOfVolunteers	Number of volunteers that work on the project	
latitude	Latitude of Facility	
longitude	Longitude of Facility	

Figure 4.03 The portion of the data entry form in which project data is entered

Since the focal point of the database is the project and per sponsor's request, it was important that we developed a standard convention for generating the structure of the primary key of that table. This went through several iterations.

At first, we considered using the UN country naming convention (KEN for Kenya, PAN for Panama) concatenated with the month and the last digits of the year in which the project began. We would often default on referring to projects in terms of the country served and the month and year the project began on site, so concatenating those concepts to the UN convention made sense to us. However, we decided against this idea after Gar Clarke indicated that we should avoid alphabetic characters as much as possible in order to sort, filter, and traversing the data more easily. Research also indicates that numeric codes minimize disk space, which allows the code to run more efficiently (personal correspondence, Shubbi Taneja, Assistant Teaching Professor of Computer Science WPI).

Our second consideration was to create an algorithm that would hash our original convention and convert it into a numeric code. After a few program designs, we decided to scrap the idea because maintaining uniqueness in the identifier would require them to be very long and difficult to keep track of. It would also require more automation and integration of other programming languages that proved to be outside the scope of our project.

Gar Clarke also advised us against using the date as part of the naming convention in case the project was ongoing or the standard for cataloging dates changed along the lifespan of the database, rendering the original data incorrect (i.e., if WEFTA decides to record the month that project development began instead of date of arrival at the project site).

Eventually we discovered that phone prefix codes were unique by country, and would make a useful numeric replacement for our UN country code. We also concurrently decided that keys for the other tables should be very simple in order to automate the process of data insertion as much as possible. SQL has the capability to generate subsequent numbers for each new datapoint, so we decided to make this new number the primary key for every other table. This number is called a sequence number in SQL.

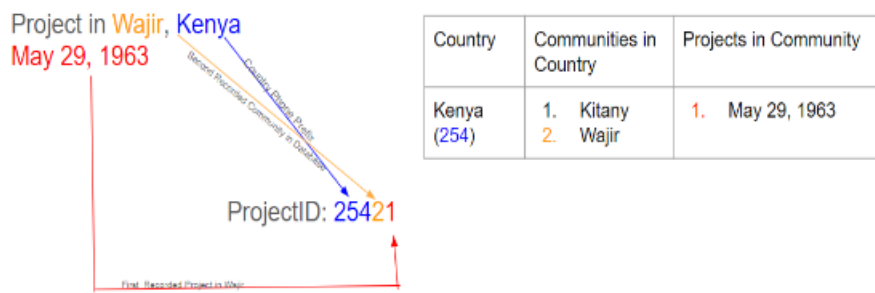


Figure 4.04 Color coded diagram demonstrating ProjectID generation

Because John DiRuggiero indicated that most projects are identified by the community served, we concatenated the sequence number primary key of the community table to the phone code. We were also reminded that multiple projects can serve the same community, so we concatenated a new sequence value to represent the order in which projects are cataloged. For example, if Kitany, Kenya is the first recorded community and the first recorded project in Kitany was in December 1992, the project code would be 25411 where 254 is the phone code of Kenya. If there were to be another Kitany project in May 2006, that project's code would be 25412. If a project were to take place in 2015 and serve Wajir, Kenya, the code would be 25421, as this project takes place in Kenya (phone code 254) and is the first time that the second recorded community was served. A project in January 2016 that serves Herrera, Panama would have code 50731 as Panama has phone code 507, and this is the first time that the third community in the database has been served. Figure 4.04 is part of our documentation, which emphasizes how a project ID may be generated.

While we initially proposed a manual for user training, we decided to instead compile all of our supplementary and explanatory material for the purpose of supporting users. Per Gar



Clarke’s advice, one of these items was a data dictionary that would list every datapoint organized by table, as well as other information about the datapoint that a user may need to know. Our original draft of the data dictionary included only data point, data type, what the value considers, and other notes, typically noting foreign and primary keys. This was later expanded to include whether or not the value was unique and the default value. This draft was retooled after a conversation about industry standard with Gar Clarke. Following Gar Clarke’s example, we also included new columns such as size, a table explaining numeric codes (i.e., 0 for poor, 2 for good), and a change log. Figure 4.05 is a screenshot of the project table of the data dictionary and demonstrates the other tabs that a user may click on to fill out different tables.

Field Name	Data Type	Size	Description	Nulls	Valid Values / Code Table	Default Value	Unique	Foreign Key	Source	Comments
projectID	Integer	15,0	Primary key made up of country phone prefix and community sequence number	No			Yes	No		Primary Key; references to Projects and Sponsors, Reported On, Pictures, Project Contacts, ProjectPartners
kpiID	Integer	10,0	Sequence Number that identifies KPI table	No			Yes	KPIs		
countrySubRegion	String	65	Name of the sub region	No			No	No		ex. Afar Region
district	String	65	Name of district served	No			No	No		ex. Asayita
communityID	Integer	10,0	Reference to Community Table	No			No	Community		
countryID	Integer	10,0	Reference to Country Table	No			No	Country		
contactID	Integer	10,0	Reference to Contact Table	No			No	Contact		
financialsID	Integer	10,0	Reference to Financials Table	No			Yes	Financials		
beneficiariesID	Integer	10,0	Reference to Beneficiaries Table	No			Yes	Beneficiaries		
projectType	String	20	Type of project	No			No	No		
projectDescription Link	String	30	Link to free form written report on project	No			Yes	No		
washConcernsLink	String	30	Link to document with Water, Sanitation, and Hygiene Concerns	Yes			Yes	No		
proposalLink	String	30	Link to proposal document	No			Yes	No		
			Link to document with							

Figure 4.05 Project table of data dictionary

The original data dictionary draft with only datapoint, type, and description was reformatted with each table in a new Excel tab to be the data entry form. We made this decision as the separate tabs and table of contents makes it easier for a user to navigate the tables and fill out only what is relevant to them. We wanted to make sure that there were as few columns as possible in the data entry form in order to prevent overwhelming the user.

One of our original deliverables was to be a Survey123 data entry form that could be converted into a comma separated value (CSV) file to be imported and visualized in Tableau or ArcGIS. However, this came to be an issue as our model required several tables. Since a Survey123 response could only be pushed to one table, it would require the user to fill out several surveys. We developed the entry form under the assumption that users could fill out the form, then John DiRuggiero could insert data into the SQL database based on the form entries.

The form, as it stands, is just a model, so John has the flexibility to interpret the model in any way he sees fit. The same goes for our database model.

Since the data dictionary and schema model were not developed in a way that was specific to any programming language, the general format can be fully realized in a variety of ways. One of John's suggestions was to develop interconnected files in Egnyte rather than a database. This model could be converted into a file system as foreign keys are synonymous with folder pointers and UNC paths.

#### 4.4 Needs Assessment Survey

As mentioned previously, WEFTA is a small organization, and we only got eight responses on our needs assessment survey. This was out of a total of nine people whom the survey was sent out to by WEFTA staff. The plan originally was to separate respondents into three categories based upon the three major roles at WEFTA. Those being board member, staff, and volunteer. The low number of respondents meant that separating the data this way was not feasible because the results would be statistically insignificant. We decided to group all the respondents together regardless of position so that the survey results would carry more weight. Once the data was collected, we needed to pick the top ten KPIs with the highest percentage of "most important" responses. This was because we determined that no more than ten KPIs could be displayed, as the human short-term memory can process  $7 \pm 3$  components at once. The KPIs with the highest "most important" selection were: project description, photos, total population, number of healthcare facilities, project cost, coordinator contact information, partner/community contact information, project type, community engagement, and total donations. See Figure 4.06 for a graphical representation of the breakdown of the top responses.

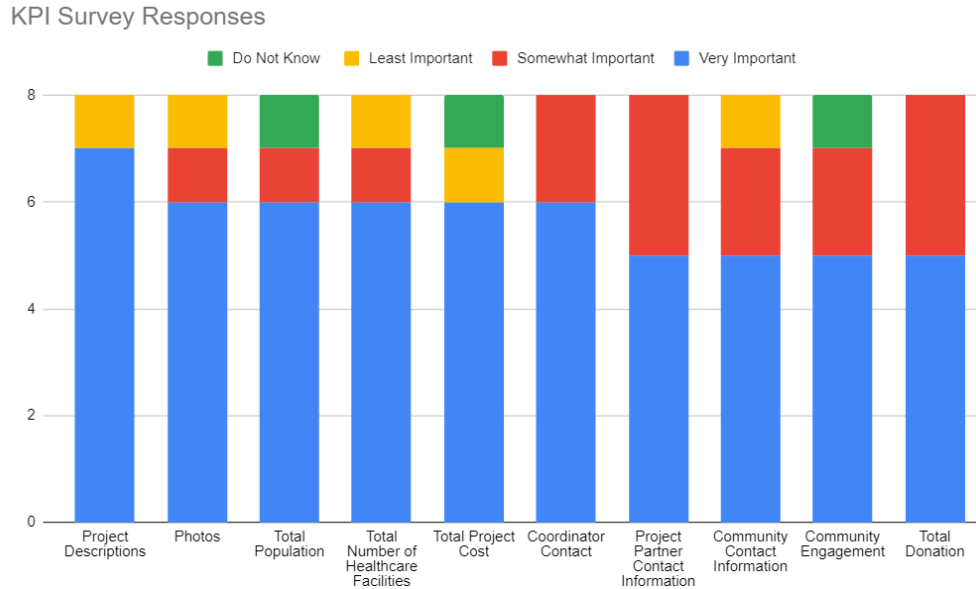


Figure 4.06 Chart of KPIs with the highest amount of “Very Important” responses

Another component of our survey was to ask respondents if they would be willing to give contact information to participate in future interviews and deliverable testing. To keep the responses from the needs assessment survey anonymous, we linked a separate survey created in Survey123 at the conclusion of the needs assessment survey to collect respondents' contact information. A high number of respondents did indicate that they would be willing to participate in future interviews and gave contact information; however, only one respondent responded to our email for an interview and deliverable testing.

#### 4.5 Dashboard

Since we discovered there were not enough responses on the survey to warrant the creation of multiple dashboards with meaningful content, we created a single overarching administrative dashboard and another financial dashboard to organize WEFTA data. The first is intended to be used by all members of WEFTA while the second is intended to only be used by members who need to have access to the financials, as determined by WEFTA. These were created in both ArcGIS and Tableau to determine software effectiveness in terms of aesthetics, capability, and ease of use. Both dashboards reflected the same KPIs and information. The KPIs included in these dashboards were determined based on the results of the needs assessment

survey. The top ten most important KPIs were selected to be displayed in the general dashboards and all seven of the financial KPIs were included in the financial dashboards.

WEFTA does not have a specific dashboard for project overviews or financials and instead has an Excel spreadsheet containing all of their project data. After discussion with our sponsor, it was found that WEFTA would prefer to keep everything, including a theoretical dashboard in ArcGIS, as that is the software they are most familiar with.

ArcGIS has very powerful spatial analysis tools which is a major benefit of the software. However, it is lacking in some of its other capabilities which makes it somewhat ‘niche’ as it is very targeted towards geospatial data. Once we began to use ArcGIS Dashboard, we realized that it is very intuitive for new users. It had a relatively quick initial learning curve, but one reason for this was because most of its data visualization tools were simple and straightforward. One feature of ArcGIS Dashboard that allowed for more advanced data visualization methods is the embedded content feature. The dashboard creator could utilize optimal data visualization methods from other software and embed them into the ArcGIS dashboard. The primary drawback of this method is that the dashboard is then pulling from different software and errors become far more likely to occur. While ArcGIS Dashboard had many other useful features such as filtering modules based on the map, these features were not unique to the software and were also available in other dashboard programs. We were able to create adequate dashboards that visualized the KPIs well and fulfilled all the requirements of a satisfactory dashboard. Since ArcGIS Dashboard’s spatial analysis tools are not utilized, the software fulfills WEFTA’s needs at a basic level but does not go beyond that.

Since the visualization methods available in ArcGIS were not ideal for the KPIs we were attempting to display, we began to explore other dashboard software, focusing particularly on the industry standard Tableau Public. Tableau is primarily used for data visualization, and is marketed as a business intelligence and analytic software. Unlike ArcGIS, Tableau’s Geospatial mapping component is not the central focus of the software, but it still has expanded mapping capabilities. It is able to layer multiple types of geospatial data and relate that geospatial data to non-locational values, much like how WEFTA uses ArcGIS to achieve the same goal. Tableau does excel in its data visualization methods, containing about 33 base visualization methods,

compared to the 11 ArcGIS has. It also has the Viz-in-Viz option, wherein one visualization can trigger and filter another (Figure 4.07).

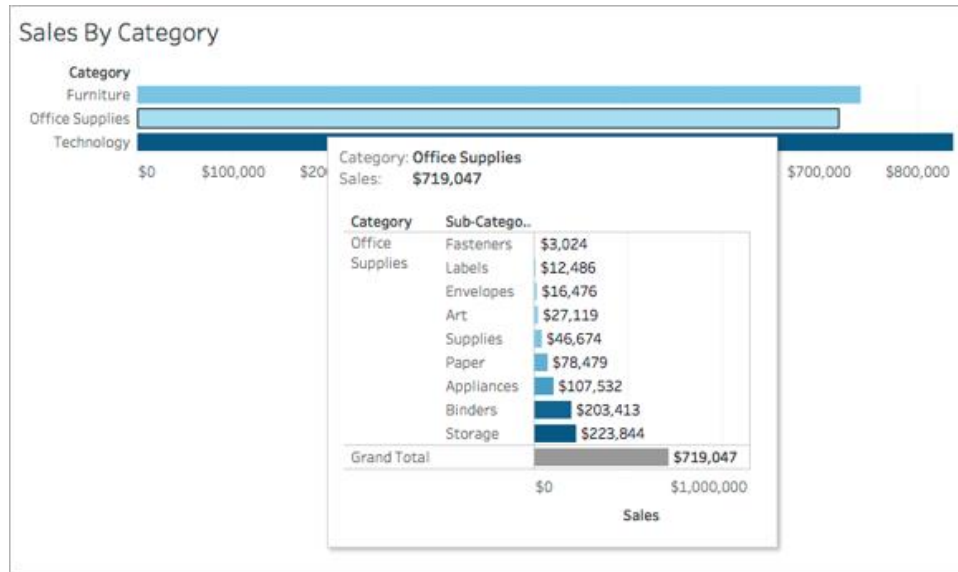


Figure 4.07 Example of Tableau Viz-in-Viz feature

This allows for layered visualization that would be well tailored to the data and accompanying notes system WEFTA currently uses. This also addresses the issue of vagueness of score details when displaying likhert scales, which is how the data is currently oriented. Additionally, Tableau has third-party plugin capabilities, so software that is not already built into Tableau can be supplemented. The plugins that would be relevant to WEFTA would likely be the R-plugin, which is primarily used for advanced statistical analysis, as well as automatic data generation. We designed the dashboard based on general design concepts, acknowledging that the Geospatial Map was the most important focal point. Therefore, the geospatial map was put in the top left corner, which is where the viewer looks when first accessing the dashboard (Figure 4.08). We also used the built-in Tableau colorblind palette when designing dashboard aesthetics. The KPI displayed were sized in order of importance from the KPI Survey. We included hyperlinks to additional project files as well as the contact information of circuit riders and coordinator contacts.

After we created the preliminary dashboards, we conducted an interview with country coordinator Paul Kennedy to get his feedback on the dashboards. Some of his critiques were easy to fix (e.g., text is too small, words are cut off, etc.) We primarily focused on his main critique

that the dashboard was not intuitive, and it is not immediately clear to a user that the map can control and change the data displayed in a dashboard. To address this problem, we considered adding a splash page or text that would indicate the dashboard was interactive, but decided not to do this because it would take up too much screen space, and there were too many interactive components to fit into a single text box. Instead, we opted to create a dashboard user training feature that would encapsulate all of the interactive features of the tableau dashboard, as seen in Appendix C. The final dashboard, post-revisions is shown below (Figure 4.08):

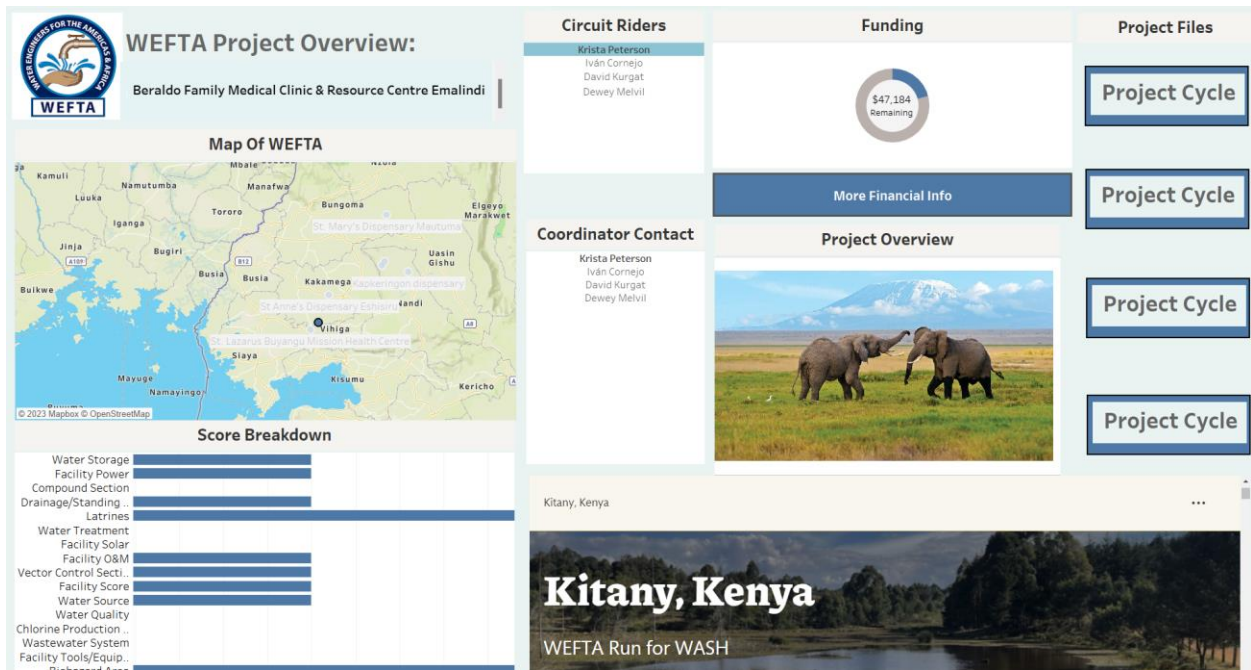


Figure 4.08 Final Tableau Overview Dashboard

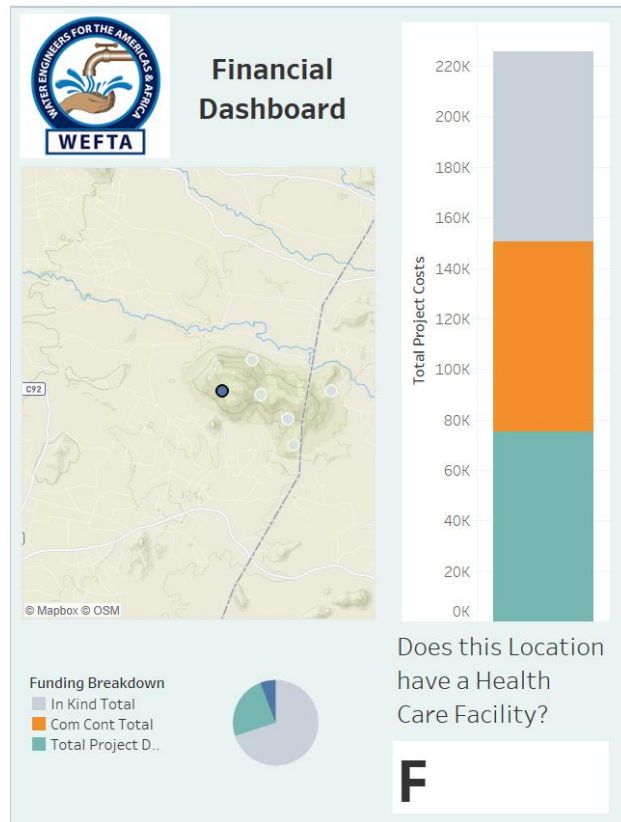


Figure. 4.09 Final Tableau Financial Dashboard

For the ArcGIS dashboard (Figure 4.10), Mr. Kennedy’s main critiques were that he could not initially tell what parts were interactive, and that it was hard to distinguish between different modules. Similar to the steps taken with the Tableau dashboard, to fix his first comment we created a section in the user training document that covers the functionalities of the ArcGIS dashboard and is available in Appendix C. However, Mr. Kennedy’s critique about distinguishing the different dashboard modules was more difficult to fix. The ArcGIS dashboard is set up in a way that allows both the general and financial dashboard to be available in two separate tabs in the same dashboard file. The only way for this to work is for all of the modules in one dashboard to be contained in a group, but grouping the elements removes the dividers between elements. Ultimately, we opted to keep the different tabs and not make a change to this both because John DiRuggiero requested having multiple tabs, and since we were only able to interview one person, we are unsure if others would have the same complaint. However, if it became a prevalent issue in the future it could be quickly remedied by simply ungrouping the elements in the dashboard and having two separate dashboards instead of two tabs.

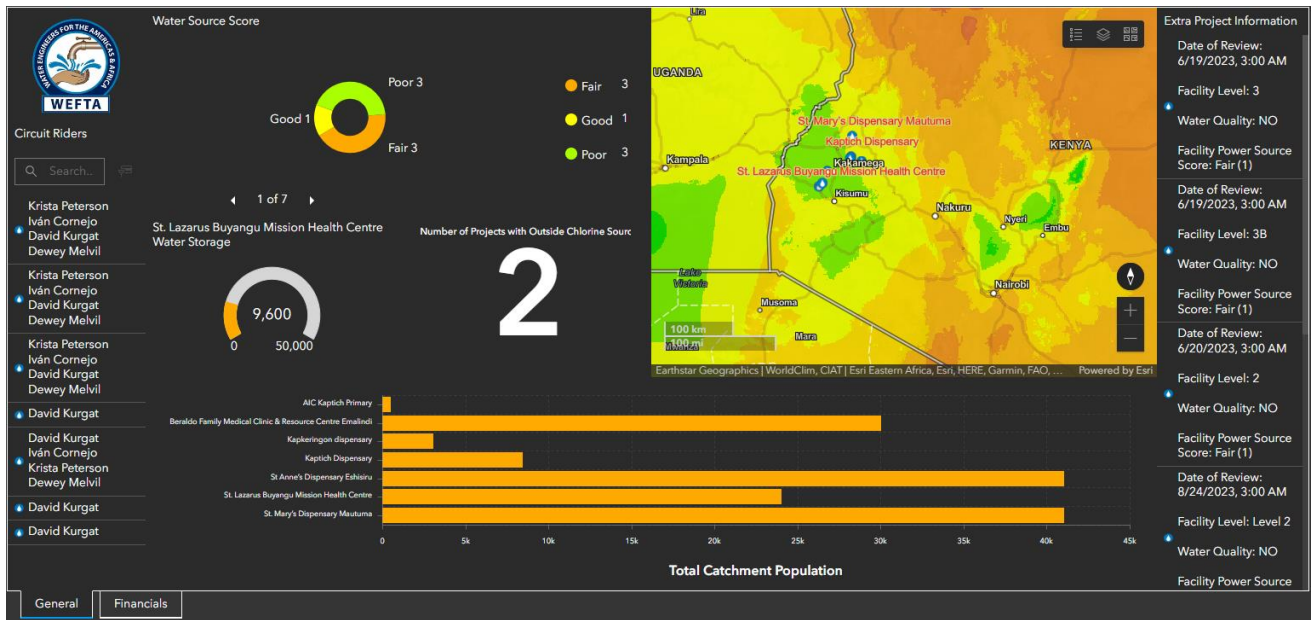


Figure 4.10 Final ArcGIS Overview Dashboard

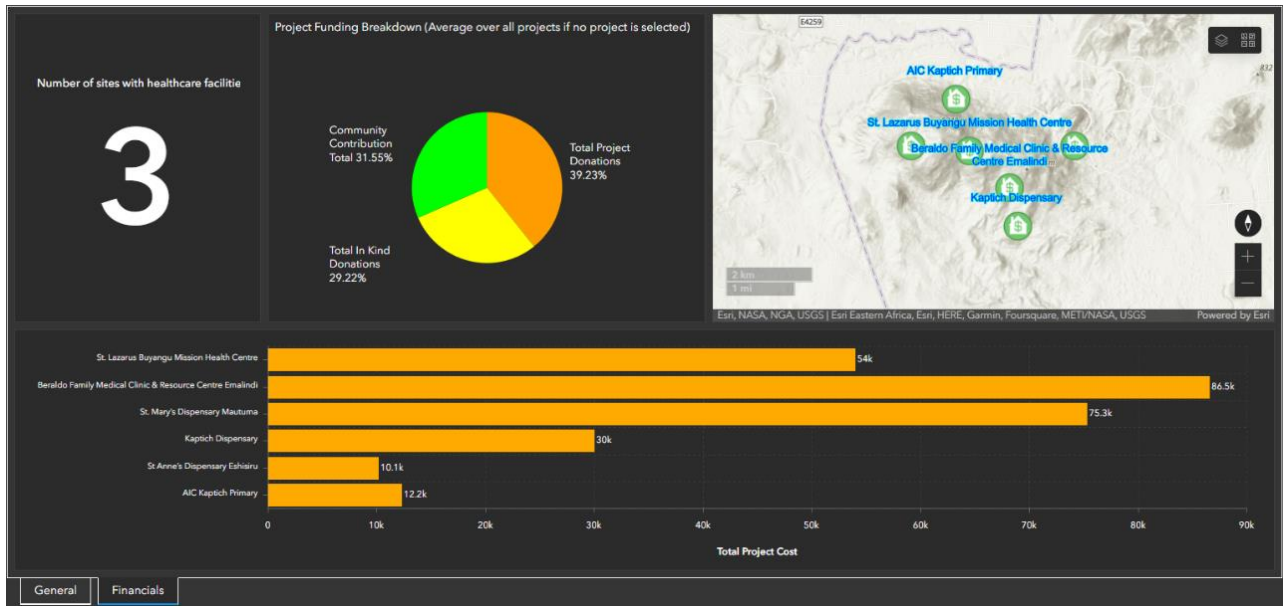


Figure 4.11 Final ArcGIS Financial Dashboard

We chose to display 7 financial KPIs and thus we were limited in visual elements to fully populate a dashboard and these dashboards have less visual content (Figure 4.09, 4.11). We have included multiple potential visualization methods of certain Financial KPIs for future WEFTA consideration in place of any unused screen-space. These financial dashboards can be used as a



visual model, and then be replaced with any financial visualizations WEFTA finds pertinent in the future.

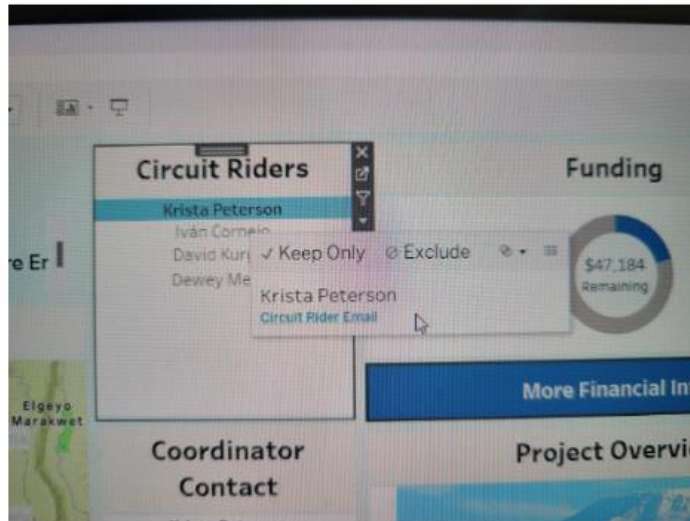
#### 4.6 Documentation and User Training

We initially planned to develop a manual to support our database, as well as a plan for supplemental stimulus training. The initial idea for this manual had multiple chapters for multiple staff titles, as well as practice puzzles to develop data entry skills. Instead, we wrote compiled explanatory content on topics that our correspondents found confusing without the additional complexities of a lengthy manual with training content and questions. We made this decision for a number of reasons. For one, WEFTA is a very small organization. Only one chapter would likely be sufficient in capturing the interest of everyone. Furthermore, we decided to reject the practice puzzles and stimulus training when we learned that developing a system for data entry would likely be outside the scope of our project. We also believed that the complexities of stimulus training would distract from the relatively straightforward nature of the dashboard and database.

We currently have four pieces of documentation. One is the data dictionary to document the parts of the database model in detail. Another is a PowerPoint demonstrating how to generate a project ID, which addresses John DiRuggiero's concerns that our proposed method was difficult to understand. The two files are descriptions of the functionality and interactivity of both dashboards. This was developed after our interview with Paul Kennedy where he mentioned that the interactive elements of the dashboard are not immediately recognizable as interactive. See Figure 4.12 for an excerpt.

2. Interactive Dashboard Components - *These are additional functional capabilities of the dashboard that must be interacted with*

- Circuit Rider Emails: To access a circuit rider's personal email, click on the circuit rider's name from the circuit rider module. A link menu with a link to their email will then appear.



- Map Interactivity: When each dot is right-clicked, all information on the dashboard will change to display only the information for that specific project center.

Figure 4.12 An excerpt from Dashboard documentation

## 5.0 Conclusions and Recommendations

WEFTA has data on several projects in upwards of 14 countries around the world, and the types of data on projects, how they have been collected, and how they are stored and managed has evolved organically since WEFTA's foundation in 2002. The lack of rigor and consistency in the types of data collected complicates data entry by staff and volunteers and inhibits easy retrieval and analysis. For this reason, we conclude that WEFTA may benefit from a standardized system of data collection for every project. We have suggested a data entry form and corresponding data dictionary that lists the data that we believe would be most useful in recording after every project.

Our proposed model may prove to be too complex to insert data into using a typical Survey123 form that gets pushed to an ArcGIS database. We recommend entering data into our data entry form first and then manually inserting the responses into an SQL database. Alternatively, WEFTA may want to consider developing a system of front-end data entry for user convenience.

The organization has also indicated an interest in storing and presenting this data in a convenient and accessible way. Our recommendation in terms of storage is to store all the future WEFTA data in a database created by database management programming language so that the most control can be exhibited over the structure. We have provided a model that we believe best organizes the data in our data dictionary.

Our sponsor has indicated a need for a naming convention for each project inserted into our model. We recommend naming by concatenating the following: country served, represented by the country's phone number prefix; a number to represent the community served; and a sequence number to represent when the project happened in terms of number of times that community was served.

WEFTA is already fully vested in the ArcGIS software and would like to develop its approach to data collection, analysis, and presentation within that environment. We explored how this might be accomplished, but we also explored the advantages of two alternative approaches with SQL as the primary database and Tableau as a visualization tool. Our

recommendation is that WEFTA consider exploring Tableau as a viable option as it could be more suitable to their needs. Furthermore, we also recommend the following regarding visualization.

(1) The use of KPIs may increase WEFTA's ability to condense metrics into quantifiable data that can be used to accurately monitor the status of projects across the globe. Based on our research, we determined that no more than ten KPIs can be displayed on each dashboard, as the human short-term memory can process  $7 \pm 3$  components at once.

(2) We recommend using the top ten most important KPIs, as determined from our needs assessment survey, in the general dashboard. These KPIs are project description, photos, total population, number of healthcare facilities, project cost, coordinator contact information, partner/community contact information, project type, community engagement, and total donations. This will ensure the most pertinent information is accessible for all members of WEFTA.

(3) We also recommend using all of the available financial KPIs within the financial dashboard. There are only seven financial KPIs currently, so they all can be displayed based upon the findings mentioned previously.

## 6.0 References

- Bishop, P. A., & Herron, R. L. (2015). Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *International journal of exercise science*, 8(3), 297–302.
- Chien, D. & Larson, D. (n.d) File structure. *MIT MECHÉ communications lab*.  
<https://mitcommlab.mit.edu/meche/commkit/file-structure/>
- Chrisman, N.R. (1999), What Does ‘GIS’ Mean?. *Transactions in GIS*, 3: 175-186.  
<https://doi.org/10.1111/1467-9671.00014>
- Darr, E. D. & Goodman, P. S. (1998). Computer-Aided Systems and Communities: Mechanisms for Organizational Learning in Distributed Environments. *MIS Quarterly*, 22(4), 417–440. <https://doi.org/10.2307/249550>
- Fong Boh, W. (2006). Mechanisms for sharing knowledge in project-based organizations. *Information and Organization* 17(1), 27-58, <https://doi.org/10.1016/j.infoandorg.2006.10.001>.
- Granke, O. (2013). Methods for database quality assessment. *Developments in Environmental Science*, 455–467. <https://doi.org/10.1016/b978-0-08-098222-9.00023-6>
- Lee, H (1995). Justifying database normalization: a cost/benefit model, *Information Processing and Management*. 31(1), 59-67. [https://doi.org/10.1016/0306-4573\(95\)80006-F](https://doi.org/10.1016/0306-4573(95)80006-F).
- IBM, *Primary and foreign keys*. InfoSphere Data Architect,  
<https://www.ibm.com/docs/en/ida/9.1.1?topic=entities-primary-foreign-keys>
- Maguire, D. J. (1991). An overview and definition of GIS. *Geographical information systems: Principles and applications*, 1(1), 9-20.
- Martin, B. O., Kolomitro, K., & Lam, T. C. M. (2014). Training Methods: A Review and Analysis. *Human Resource Development Review*, 13(1), 11–35.  
<https://doi.org/10.1177/1534484313497947>
- Meyerson, D., Kramer, R. & Weick, K., & (1996). *Swift trust and temporary groups*. SAGE Publications, Inc., <https://doi.org/10.4135/9781452243610>
- Suarez, C., Kearney, D., Davis, A., Ginwright, S. A., Benedict, I., Ghabdian, W., Ross, J., & Gottlieb, H. (2020, December 9). *Models and components of a great nonprofit dashboard - non profit news: Nonprofit quarterly*. Non Profit News | Nonprofit Quarterly.

<https://nonprofitquarterly.org/financial-management-models-of-a-great-nonprofit-dashboard/>

Tseole, Nkeka P., et al. “Barriers and Facilitators to Water, Sanitation and Hygiene (WaSH) Practices in Southern Africa: A Scoping Review.” *PLOS ONE*, vol. 17, no. 8, 2 Aug. 2022, p. e0271726, [www.ncbi.nlm.nih.gov/pmc/articles/PMC9345477/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC9345477/), <https://doi.org/10.1371/journal.pone.0271726>.

Ullman, J.D. & Widom, J. (2008). *A first course in database systems: third edition*. London, England.

UNICEF. “Water, Sanitation and Hygiene.” *UNICEF*, 2021, [www.unicef.org/wash](http://www.unicef.org/wash).

UN. “WASH – Water, Sanitation and Hygiene.” *UN-Water*, [www.unwater.org/water-facts/wash-water-sanitation-and-hygiene](http://www.unwater.org/water-facts/wash-water-sanitation-and-hygiene)

Verdinelli, S., & Scagnoli, N. I. (2013). Data Display in Qualitative Research. *International Journal of Qualitative Methods*, 12(1), 359–381.

<https://doi.org/10.1177/160940691301200117>

Walters, K., & Rodriguez, J. (2017). The Importance of Training and Development in Employee Performance and Evaluation.

WHO. “Water, Sanitation and Hygiene: Exposure.” *Www.who.int*, [www.who.int/data/gho/data/themes/topics/topic-details/GHO/water-sanitation-and-hygiene-exposure](http://www.who.int/data/gho/data/themes/topics/topic-details/GHO/water-sanitation-and-hygiene-exposure).

## **Appendix A: Needs Assessment Survey**

The KPI survey began with a preamble explaining the purpose of the survey, how it will be used, and the participants' anonymity. The preamble is as follows:

- This survey is being conducted by Worcester Polytechnic Institute students in collaboration with WEFTA to help organize and expand their dashboard capabilities. The survey is anonymous and will take approximately 2 minutes. Your response to this survey will help gauge what Key Performance Indicators (KPIs) and data should be made available for viewing and editing on a WEFTA dashboard. Not all KPIs are applicable to all projects but may be used in future projects. We will not collect any personal information that can be associated with your answers, although we will ask you for a general description of your relation to and role at WEFTA. We will ask those interested in participating in a follow-up interview for contact information to set up the interview, but this information will be collected separately to maintain the anonymity of the survey.

The survey began with a drop down selection asking participants to select the option that best describes their relation to WEFTA. Those options would include: board member, staff, volunteer, and other. If “other” is selected, then another question would pop up asking the elaborate. If “volunteer” is selected, then a question asking to describe the extent of the work they conducted will pop up. If “staff” is selected, then a question asking to describe their title and role within WEFTA will pop up. This will be the same for “board member.”

Once the background information was collected, the survey went on to ask participants to rank KPIs that were derived from information received from WEFTA. They were asked to rate KPIs from “very important,” “somewhat important,” “least important,” or “do not know.” The KPIs were grouped in the following categories: general project information, community information, finance, and technical information. Another question then asked if the respondent had other KPIs that they would like to include that were not mentioned. The final question asked if the respondent would be willing to give contact information for future interviews and testing.

## **Appendix B: Contact Information Survey**

Preamble: This survey is being conducted by Worcester Polytechnic Institute students in collaboration with WEFTA to help organize and expand their dashboard capabilities. By filling out this survey, you are agreeing to possible future interviews where you may be asked to test a prototype WEFTA dashboard.

Respondents of this survey were asked to following questions:

1. Please enter your name and position at WEFTA:
2. Preferred email address:



## Appendix C: Interview Questions

Paul Kennedy was asked the following questions during our interview:

Introduction:

Hi \*interviewee name\*, we just wanted to thank you for taking the time to meet with us and give us feedback on our Preliminary Dashboards. We will be revising our dashboards based on your responses, is okay if we quote you and your suggestions in our findings?

Tableau:

To begin we'll give you a few seconds to play with the dashboard before moving onto some basic tasks to assess functionality. Some functions will redirect to the same content. This is not an error, we are limited in data and would like to assess functionality and create a model that can be populated later.

1. First off what are your initial opinions of the aesthetics and look of the dashboard?  
Further prompting: How do you feel about the color scheme?  
The sizing of modules  
The layout/balance of modules
2. Functionality Tasks
  - a. Can you please tell me the name of a circuit rider for the Beraldo Family Medical Clinic?
  - b. Can you please tell me how much more funding the Beraldo Family Medical Clinic requires?
  - c. Based on the information provided from the dashboard, please tell me some notes on the water source.
3. Dashboard Navigation
  - a. Does the map/buttons feel intuitive?
  - b. Is navigation between dashboards intuitive?
  - c. Do you feel there is anything missing?

ArcGIS:

To continue we'll give you a few seconds to play with the dashboard before moving onto some basic tasks to assess functionality.

4. Next what are your initial opinions of the aesthetics and look of the dashboard?

Further prompting: How do you feel about the color scheme?

The sizing of modules

The layout/balance of modules

5. Functionality Tasks

- a. Can you please tell me which facility has a facility level of “3B” and a facility power source with a score of 1 or “fair”
  - b. Can you please tell me the size of Beraldo Family Medical Clinic & Resource Centre Emalindi’s water storage?
  - c. Can you please tell me the circuit riders at St. Anne’s Dispensary Eshisiru
6. Do you see how to switch between different dashboards? Is it easy and accessible?

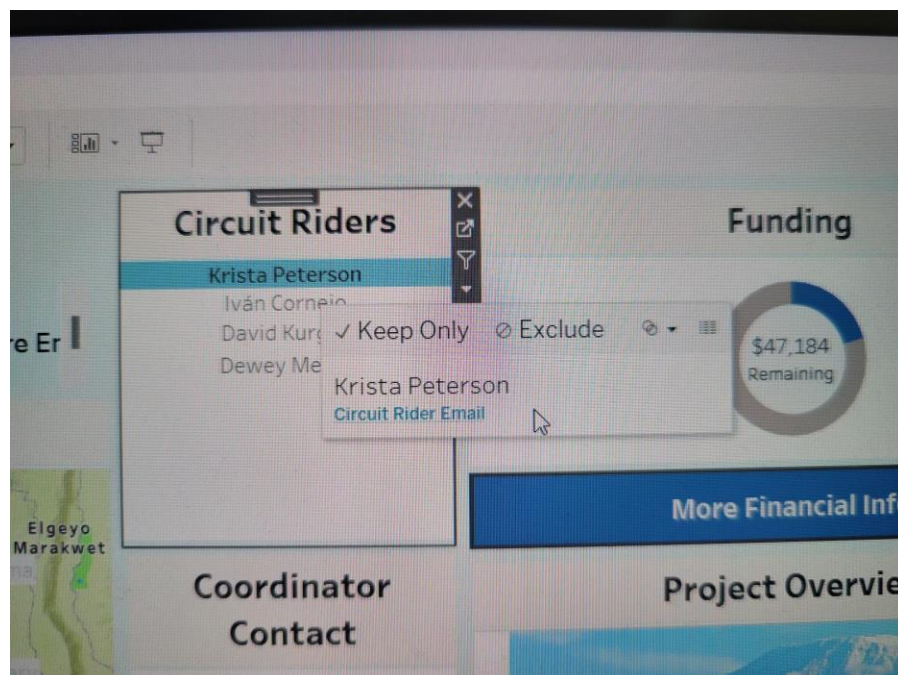
Concluding Thoughts:

Thank you for taking the time to meet with us. Your feedback has been invaluable.

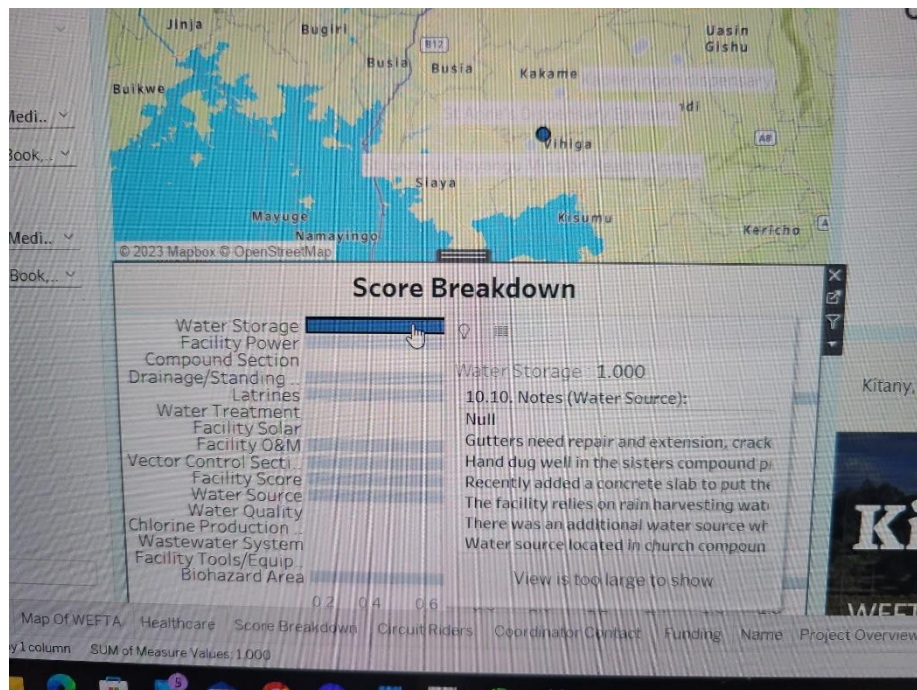
## Appendix D: Dashboard Training Manual

### Tableau:

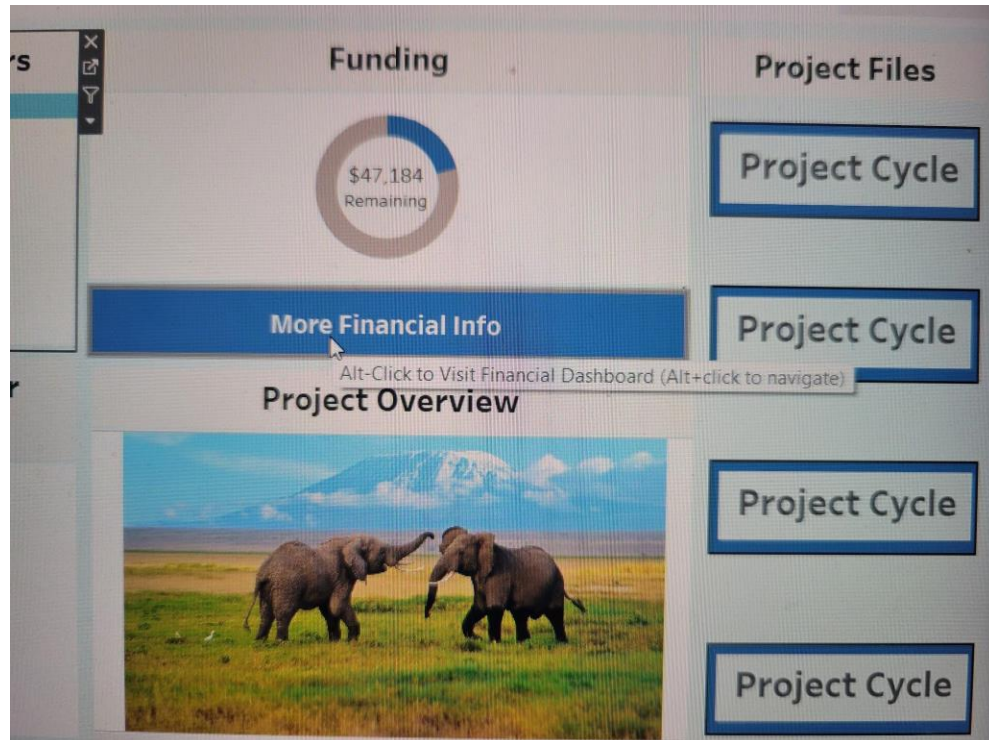
1. Dashboard Contents - *All this information is available and accessible by any staff member on the WEFTA Overview Dashboard*
  - Project Site Score Breakdown
  - Geospatial Map
  - Circuit Rider Contact Info
  - Coordinator Contact Info
  - Project Funding Breakdown
  - Project Photos
  - Project Profile Webpage
  - Additional Project Files
2. Interactive Dashboard Components - *These are additional functional capabilities of the dashboard that must be interacted with*
  - Circuit Rider Emails: To access a circuit rider's personal email, click on the circuit rider's name from the circuit rider module. A link menu with a link to their email will then appear.



- **Map Interactivity:** When each dot is right-clicked, all information on the dashboard will change to display only the information for that specific project center.
- **File Connectivity:** To access any additional project site files, click the corresponding project file button on the far right of the screen. A new tab will open with the corresponding file.
- **Score Breakdown:** To access the notes and additional comments for a project score, hover over the bar corresponding to the score you want to see comments on. When the bar is clicked on it will also re-direct to a separate page with those notes.



- **Financial Dashboard:** To view more information about funding breakdown, Alt-Click on the “more financial info” button. This will open up a new tab with the financial dashboard. NOTE: This may be restricted based on user permissions



#### Financial Dashboard Contents:

- Project cost breakdown (includes percentage funded by in-kind and other donation types when hovered on)
- Map Interactivity
- Healthcare Facility Component

#### ArcGIS:

##### 1. General Dashboard Contents

- Project Water Source Score
- Geospatial Map
  - Layer depicting total annual precipitation
- Circuit Rider Names
- Date of Review
- Facility Level
- Water Quality
- Facility Power Source Score
- Facility Water Storage
- Facility Chlorine Source
- Total Catchment Population

## 2. Financial Dashboard Contents

- Geospatial Map
- Project Funding Breakdown
- Total Project Cost
- Number of Projects with Healthcare Facilities

## 3. Interactive Dashboard Components - *These are additional functional capabilities of the dashboard that must be interacted with*

### General Dashboard:

- Circuit Rider Module Interactivity
  - The module has a search bar on the top where a user can search for a specific circuit rider
  - A given set of circuit riders can be selected to pan and zoom the map to a specific project. This will also filter all the other modules to that specific project
  - To unselect the set of circuit riders and remove the module filters, click on the selected set of circuit riders.
- Map Interactivity
  - Buttons in the top right of the map allows layers to be hidden or shown and allows the basemap to be changed.
  - All dashboard modules except for Water Source Score and Total Catchment Population can be filtered for a specific location by selecting that location on the map
  - All dashboard modules can be filtered for one or many specific locations by panning and zooming in/out on the map so the view only includes the desired locations
  - To unselect a selected map location, press the “x” button in the top left or click an arbitrary location on the map
- Water Source Chart Interactivity
  - One or multiple sections of the pie chart can be selected to filter the map and other modules for those specific projects
  - Users can click any attribute from the legend to the right of the chart to exclude that section from the view
  - To unselect or reselect anything, click the same thing that was clicked to select or unselect it
- Water Storage Interactivity

- Arrow buttons above the title of the water storage module can be used to cycle through the water storage for all projects
- The arrow buttons will disappear when filtered by a different module
- Total Catchment Population Interactivity
  - Any bar can be selected on the graph to filter all the other modules to that specific project
  - The bar can be unselected by pressing on it again
- Extra Project Information Interactivity
  - Any set of project information can be selected to pan and zoom the map to a specific project. This will also filter all the other modules to that specific project
  - To unselect the specific project and remove the module filters, click on the currently selected information again

Financial Dashboard: *To switch to the financial dashboard press the “Financials” tab at the bottom of the dashboard*

- Map Interactivity
  - See “Map Interactivity” in General Dashboard
- Total Project Cost Interactivity
  - See “Total Catchment Population Interactivity” in General Dashboard