

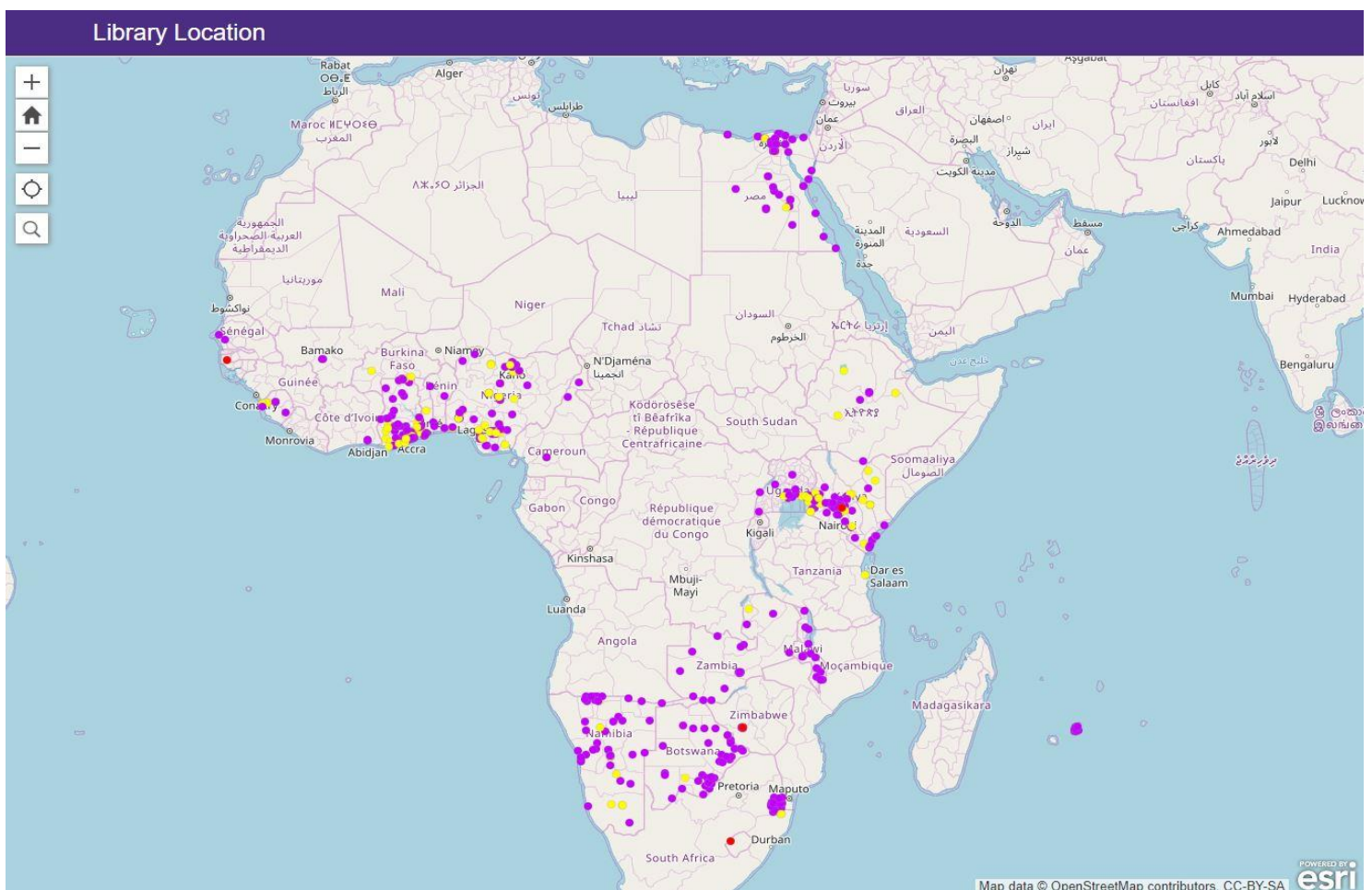
Placing trust in crowdsourced data; Using geographical information of libraries in sub-Saharan Africa

By: Mark Johnson and Amy Beller

GEOG 569 GIS Workshop

August 23rd, 2019

Report Submitted to Technology & Social Change Group (TASCHA)



Executive Summary

Information technology is still an adolescent idea in parts of the world today. Electronic connectivity along with global information is premature in parts of sub-Saharan Africa that need it the most.

In many parts of the world, libraries are where the impact of global information can be measured. Libraries in rural areas are used to access information for literacy improvement, employment improvement, and to gather communities together. All vital instruments for bringing together and sustaining healthy communities.

In rural sub-Saharan Africa, libraries there may not be noticed when looking at the global information infrastructure, due to lack of knowledge of their locations. This could be due to them being small, underfunded, or too far away from major cities to be properly equipped, or to take advantage of the internet access that is being established in many parts of sub-Saharan Africa. In order to do this, they will need to be seen and properly funded. But how can they be funded, if potential funders cannot properly locate them in the global information infrastructure?

Technology & Social Change Group (TASCHA) has partnered with African Library & Information Associations & Institutions (AfLIA) to create The Advancing Library Visibility in Africa (ALVA) project. Using ArcGIS online, these libraries can make themselves known using crowdsourced driven data. Unfortunately, with data gathered through crowdsourced methods, there can be some discrepancies. Data can have inaccuracies due to many different issues, such as misuse of the location technology, duplicates of data, and lack of technical understanding of ArcGIS online. Having these insecurities within the data can make it difficult for users to look at the crowdsourced data and know which data to trust. Implementing a measurable way for the data to be verified would greatly improve the way that crowdsourced data is examined. As well as, for ALVA to save time going through the entire data by hand making sure that it is accurate or trustworthy.

For this project, we designed a theoretical framework that would designate a system of credits based on how well each set of library location data can be verified. Using the questionnaire that is asked of all library locations, the data table, and knowledge of the system infrastructure of the crowdsourced data, we've designed a system of verification that we call trust verification credits, that can be obtained by library locations based on their verification status. Things like accurate location, full address and contact information, are just a few things that can raise or lower trust verification credits for a library location. Those trust verification credits will be allotted based off the different types of user believed to be inputting library locations, those being; random creators, trusted creators, and librarians. For each user there are a certain amount of trust verification credits each can gather. Due to the nature of the data, this is an evolving project that may grow over time.

The goal for this project is to find a way to help the public, look at the libraries crowdsourced data, and decide whether to trust the information provided. Having the ability to discern different levels of

trustworthiness or verification between library locations will greatly improve those libraries ability to be recognized, to obtain funding in the future.

Table of Contents

1. Background and Problem Statement	1
1.1. Background	1
1.2. Problem Statement and Project Goal	1
2. System Resource Requirements	3
2.1. Data Source	3
2.1.1. Database	4
2.1.2. Database Quality Description	5
2.2. Hardware and Software	6
3. Business Case Evaluation	7
3.1. Benefits	7
3.2. Costs	9
3.2.1. Capital Costs	9
3.2.2. Operating Costs	10
3.3. Benefit Cost Analysis	11
4. Data Development	13
5. Workflow Implementation	14
6. Results	16
6.1. Data Table Alteration	18
6.2. Trust Verification Credits	20
6.2.1. Trust Verification Credits Framework	21
6.3. Login Program	23
7. Conclusions and Recommendations	23
7.1. Conclusions	23
7.2. Recommendations	24
8. References	25
9. Technical Appendices	25
9.1. Appendix A: Contacts	25
9.2. Appendix B: Blogs	25
9.3. Appendix C: Programs and Software	25

1. Background and Problem Statement

1.1 Background

Technology & Social Change Group (TASCHA) from The University of Washington Information School, and African Library & Information Associations & Institutions (AfLIA) have created a joint research project called Advancing Library Visibility in Africa (ALVA). AfLIA's goal is to bring information technology to libraries in rural parts of Africa. Partnering with TASCHA, they want to get rural libraries in sub-Saharan in Africa noticed to aid in that goal. Libraries are important in these areas because they promote literacy and learning in underprivileged groups, such as women and at-risk youth (Osuigwe et. al. 2018).

In TASCHA's project *Visualizing Crowdsourced Geographic Information with Unknown Accuracy* they want to take on this investigation in order to make it easier for the data of the rural library locations to be made accessible to the public. Their initial step in bringing information technology to rural libraries is to make them known in the first place. ALVA has created an online ArcGIS map that each of their in-country Champions have been given the link to, in order to distribute to their local libraries. With this link, many libraries have entered their information and locations into the ArcGIS online map. In order to input their location, they must follow the link, fill out a questionnaire and locate their library on the map and then submit for approval which is done by ALVA.

Using the online ArcGIS map format, ALVA can gather crowdsourced data of the locations of many libraries that otherwise may go unnoticed. Crowdsourced data is a wonderful way to gather lots of data in a shorter span of time and that might otherwise be uncirculated. However, the negative to this data gathering technique is that crowdsourced data can face several inaccuracies. Those being, things such as locations not being properly pinpointed into the map or data being misunderstood.

As of now, there are a few cases where you can find several instances of one library located in multiple locations. Or maybe several different libraries with the same name, that have just not been reviewed yet. Since the data is in a constant state of growth, having a way to examine the data gathered and deciding which is the most trustworthy is needed for a project of this magnitude to progress.

1.2. Problem Statement and Project Goal

There is a way to ensure data accuracy, especially through crowdsourced, by using the most important data source – the public. By creating a system to ensure the reliability and accuracy of the public input, we can then use this huge data source to ensure the accuracy of the sources of data entered into the library locations map. With this thought process, we have designed a system of verification, that we call trust (or verification) credits, that will be allotted to the library entered based on how trustworthy the data is. Data is inputted into the online ArcGIS Map system and verified, at first for approval by ALVA, then through our trust credit system, in theory. As the system grows and other data sources are

entered into the system, these library locations sources will then be used to verify new data locations as a form of support, thus verifying the reliability of the new data sources. All of this can be uploaded into the data table that is being used by ALVA now.

Our goal for this project is to find a way to help the public look at the libraries crowdsourced data and decide whether to trust the information. We have made it our objective to create a scoring system that assigns credits to the library's uploader based off a series of qualifications and to aid in the expansion of a more comprehensive way to ask questions in the questionnaire that is asked of the libraries when they submit a location. Inputs will be given trust verification credits for varying things, such as properly filling out the questionnaire, correct (or seemingly correct) pinpoint locations, and full address, to just name a few. In creating/implementing this credit system, we will design a framework that uses the data table created from the crowdsourced data to assign those credits. In doing so, we will need to clean up the current data table, as there is a lot of unused fields that do not pertain to the current data set anymore due to updates and general evolution of the data over time. We would then change the data table to mirror the information that we will gather trust verification credits for. This will help any new incoming data, but for older data there will have to be updates from the uploaders/creators of those library locations.

Being that our goal for this project was to create a way for users to look at the data provided and make an informed choice on which data is more trustworthy, we thought it would make more sense to create categories of input users based on their accumulation of trust verification credits. In order to distribute trust verification credits accordingly, we designated the types of users into three categories; random creator, trusted creator, and librarian. Within each category are subcategories of trust verification credits that the user can be allotted. In the ending result, the accumulation of credits will help ALVA and those who use the data to make an informed choice on which data can be trusted based on the verification process.

2. System Resource Requirements

2.1. Data Source

All data for this project was obtained through crowdsourced data gathered from an online ArcGIS map created by the joint research project, Advancing Library Visibility in Africa (ALVA) done by Technology & Social Change Group (TASCHA) and African Library & Information Associations & Institutions (AFLIA). The data was entered by users or librarians in sub-Saharan Africa who were given the link to the ArcGIS website by ALVA's in-country Champions. These are the people who are in the district of those libraries and in charge of making sure that they put their locations into the online map. Librarians/users will follow the link given to them and fill out the questionnaire shown below made by ALVA.

< Library

Click the map to locate your library. / Cliquez sur la carte pour localiser votre bibliothèque. / Clique no mapa para localizar sua biblioteca. / انقر فوق الخريطة لتحديد موقع المكتبة.

Click the map to draw the location.

Enter an address to search

Details / Détails / Detalhes / تفاصيل

Name of Library / Nom de la bibliothèque / Nome da biblioteca / اسم المكتبة

Country / Pays / País / بلد

Type of library / Type de bibliothèque / Tipo de biblioteca / نوع المكتبة

Other Library Type / Autre type de bibliothèque / Outro tipo de biblioteca / نوع مكتبة اخرى

Street Address / Adresse physique / Endereço da rua / عنوان الشارع

Name of community or town / Nom de la communauté ou de la ville / Nome da comunidade ou cidade / اسم المجتمع أو البلدة

Library Contact Name / Nom du contact de la bibliothèque / Nome do contato da biblioteca / اسم جهة اتصال المكتبة

Email address of library contact / Dirección de correo electrónico de contacto de la biblioteca / Adresse électronique de contact de la bibliothèque / عنوان البريد الإلكتروني لاتصال

Your name / Tu nombre / Seu nome / اسمك

Your email / Tu correo electrónico / Seu endereço de email / عنوان بريدك الإلكتروني

Attach a photo of the library. / Joindre une photo de la bibliothèque. / Anexe uma foto da biblioteca. / ارفق صورة من المكتبة.

To the right of the questions is a map in the ArcGIS online program, where the user will select their location. Each question in the library information detail is then entered in the data table on ALVA's website. When questions are not filled in, they are left with a <NULL> in the data table.

Being that the data is collected via crowdsourced, it can be done by library employees or by random people, essential anyone with the link. For data accuracy after data input, reviewers go through the data table and try to verify the inputs looking at various sources and occasionally reaching out to the data provider, if an email was given. Currently, ALVA's validation process for the crowdsourced inputs involves using sources such as satellite imagery, OSM, Google Maps/ Street View, and social media to improve the validity of the likelihood that there is a library in the location inputted into the database.

When a location reviewed is deemed inaccurate, the field is labeled rejected and on the map the point color is changed to red. Originally it was intended to be deleted from the map, but users were under the impression that their inputs were not being uploaded so they kept submitting the same information, so now rejected reviews are remaining visible on the map. There are three colors for the dots currently on the map, purple; approved, yellow; pending, red; rejected. As more locations are approved more will become purple. From these approved points we will build our system of trust verification credits from. This gives the user security when deciding which data is accurate to use.

2.1.1. Database

Filename/Location: LibraryLocation -

<https://uw.maps.arcgis.com/home/item.html?id=e9b07c08ea6c47abaf922274acfb5e2d&view=table#data>

The data file is a table with 25 fields, containing information about libraries in sub-Saharan Africa. This table includes information about library name, location, data upload information, point contributor, and point verification.

Spatial object type - point

LibraryLocation – point feature class

<https://uw.maps.arcgis.com/apps/CrowdsourceReporter/index.html?appid=8a4e7f85628f405c81d8cbbc6fee110>

<https://uw.maps.arcgis.com/home/item.html?id=e9b07c08ea6c47abaf922274acfb5e2d&view=list#overview>

CROWDSOURCING AFRICAN LIBRARIES

NAME	DATA TYPE	NULLABLE	USER INPUT	TO BE DELETED	GEN DESCRIPTION
OBJECTID	OBJID	N			Object ID
LibraryNa	STR	N		Y	Name of Library - now replaced with LibName
CountryNa	STR	N	Y		Name of Country
LibraryTy	STR	Y	Y		Type of Library (Academic, Community, Public, National, School, Other)
OtherTy	STR	Y	Y	Y	User can indicate what other type of library is, if choosing other in LibraryTy
StreetAd	STR	Y	Y	Y	Address of Library - now replaced with StrAddr
CommOrSub	STR	Y	Y		Name of Community or Town library located in
LibContact	STR	Y	Y		Head librarian to contact at library
LibContactMeth	STR	Y	Y		Librarians email address
RespondNa	STR	Y	Y	Y	Contributor name - same as Creator but bad name
RespondEmail	STR	Y	Y		Contributor email - should be renamed to CreatorEmail
UserID	STR	Y			Unique identification for contributor - not used
Reviewed	STR	Y		Y	(SHOULD BE BOOLEAN) - Indicates if location was reviewed - now replaced by ReviewStatus
ReviewerCO	STR	Y			Reviewer Comments - will need to be adjusted for multiple reviews
Date	DATE	Y	Y		Date of Creation of input
GlobalID	STR	N			Unique ID for library created on creation
CreationDate	DATE	Y			Date of Creation of input - redundant
Creator	STR	Y			Creator Name - Same as RespondNa
EditDate	DATE	Y			Date of most recent edit
Editor	STR	Y			Editor name
DataSource	STR	Y			Source of data
LibName	STR	Y			Name of library - input by contributor
StrAddr	STR	Y			Address of library - input by contributor
ReviewStatus	STR	Y			Indicated if location has been reviewed

2.1.2 Database Quality Description

The current table done by ALVA needs slight restructuring, as there are unused fields and some multiple fields for the same data. The crowdsourced data will hopefully be from librarians and/or their employees, so we expect to have at least a sense of reliability. The major concern is that even though this program is primarily going to be used by librarians, it does not necessarily mean that each librarian will be educated to the same level, or even have the same standard of resources available to them. The concern is not necessarily that any data inputted could be corrupt, but that it might not be 100% trustworthy.

That is the whole reason we have designed a theory of a co-user verification process as part of the trust credit validation system. This will require a login process; through this process each user will build a user status which will ultimately create a community that can help determine and filter who can be trusted for verification purposes. As well as users helping verify others, saving time for ALVA, it will also gain them credits for doing that verification. In the end this will help make this platform self-sufficient and rely less on the independent verification process that ALVA uses now. After all, who can be better filters about libraries than librarians themselves?

Eventually, anyone who logs into the system will be able to look up a library, find exactly where it is, and what type of library it is. Or for analysis purposes they could investigate various locations and find how many locations have certain types of libraries. Perhaps even how far a potential library customer may have to travel in order to find a library for a type of program being offered.

2.2. Hardware and Software

For this project the hardware and software are simple and easy to obtain. Being crowdsourced data, there isn't much in the way of hardware needed other than the necessities to gain access to the internet. As far as the software is concerned, the data is gathered from ArcGIS online and is the only software needed.

As a librarian/user enters data they would need technology that has access to the internet, be it a smartphone or a computer to input the data. For software they would need the link to the online ArcGIS map to input their location.

For a member of ALVA, they would need the same, in addition an ArcGIS account from Esri to be able to format and compile their gathered data.

For our project, we used personal laptops and our Esri accounts to download and examine the data tables gathered from the crowdsourced data on the online ArcGIS Map. As well as ArcGIS Pro and Python for the design of the login system for the verification system.

3. Business Case Evaluation

Currently the system created by TASCHA for their data gathering is basic, but slightly misleading. Anyone with the link can input data. For the team responsible for verifying the data, there is not currently a standard in which to check the accuracy, besides web and social media checks. Our recommendations of data table restructuring, and trust verification credits can help with that. But as with any benefits, it isn't without its costs. We believe that our project goal will help in decreasing costs for ALVA's Project, saving time and money.

3.1. Benefits

By restructuring the current table schema, the process will be more direct and less misleading for the user. This will ensure data is not added simply to 'fill in a blank'. As well as having a select class of trusted users being held responsible for many of the data verification, this will ensure that data is verified more quickly and there will not be a need for a ground verification crew. This will give the verification team of ALVA more time to focus on specific locations or maybe to double check possible locations that may be pending approval.

Our theoretical system where users can log in and gain credits for becoming trustworthy, may eventually invite potential investors to explore the system and look at popular libraries for potential investment or donations.

There could be a possibility, that the more users who utilize this system, once it becomes a stable and trustworthy system, than more librarians and other types of investors can use it for different analysis purposes. This could include determining distance between libraries vs population with the help of utilizing another map application such as google maps.

The Antenucci et al. GIS: Guide to the Technology (1991 pp 66- 72) typology lists 5 types of benefits: Type 1: Currently there are efficiencies in practice, the biggest being that the system is quite simple to work with. Once the user inputs their location and fills out the questions for the ArcGIS online map set up by ALVA, that data is then sent to a corresponding data table. The entire network is contained on the basic table, which holds all the data that is collected for locations. The users who input the data, do not yet have access to the data table. The concept is simple to work with, but labor intensive as each new location must be individually checked by ALVA before approval. If any problems do arise, with the locations, photos, or information entered, then the library data must be cross referenced with another source, such as Google Images or by social media. With our project, it will save time and labor for ALVA, estimated around 15-20 mins. Since the process of checking will be less frequent, with the incorporation of an altered design for the questionnaire and the data table, as well as the giving the users an informed way to make decisions on which library locations are the most trustworthy.

Type 2: Currently anyone can look at the ArcGIS Map application and place the location of a new library anywhere they want. When users enter a new library, a series of questions are asked that they are prompted to fill in. The collection of questions asked, as well as the order they are asked in, are somewhat misleading. Users are also required to click on the map to pinpoint exactly where this library is located and when the map isn't zoomed in correctly this can warp the data since there is no particular range or map scale recommended in the instructions for the user to go off of. By restructuring the table, the set of questions being asked, reworking pinpointing the location, as well as creating a system of co-user data verification; the plan is to make the data being placed into the map more reliable on entry, as well as more quickly verified by a system of active users. This will give the current work force the opportunity to be more specific in field checking, more crucial locations, as well as potentially focusing on other aspects of the network.

Type 3: Once this system of trust verification credits is initialized and proven to be stable, there are many potential or unpredictable events that could come from this network. This network could be created to invite potentially interested investors to libraries that may not have been known. Because of this network, knowing the library location and that the presented information can be trusted, these investors may potentially reach out to a location they did not realize was present before. There could potentially be an influx of beneficial funding with the help of this accuracy verifying system. Librarians might use this system to find a way to communicate between locations to determine a way to share resources, which may not currently be established. With the possibilities of using this network, as well as others (such as google earth or yahoo maps), it may help the establishment of new libraries in areas that do not have the resources currently present.

Type 4: One intangible advantage from this network is that librarians could use this network to locate other libraries. In the process, they could look at other information within the table including, the type of library and who is the primary contact. Through this, the librarians may find themselves exchanging information with other librarians and learning and improving their own operations. As the network becomes more stable and trustworthy, say for example when a library users go to one library looking for resources, maybe that library doesn't have what they need, the library then could contact another library nearby using the network, and quickly direct the customer to where they may find this resource. Another intangible benefit that may arise is with funding. When potential investors are looking for possible libraries to fund, they may find them through this network, using the resources to find a library more specifically suited for their philanthropic goals.

Type 5: There will be no sale of information as this is meant as an open data source, primarily for librarians and potential investors to use. One of the major goals to this platform is the promotion of investment opportunities of libraries and education in sub-Saharan Africa. Ultimately the end

goal would be for knowledge of those libraries existence throughout sub-Saharan Africa be available to support and promote a solid infrastructure.

3.2. Cost

There are a few costs to the implementation of our project, those being labor costs, the creation of the trust credit systems and data table design.

The Antenucci et al. GIS: Guide to the Technology (1991 pp. 72-78) cost typology consists of:

3.2.1. Capital Costs

1) Database

Being that the data is purely crowdsourced derived, we should account for the collective hours of each person who entered their library into the online ArcGIS map. As well as data table subsequently created from this data. Then the table of trust verification credits that was created for the better trustworthiness of this data.

In amount of time used for these activities, we would estimate that for every user (so far) who has input a location on the map in the ArcMap would be about 7 minutes for each entry, with hundreds of entries resulting in anywhere from 50-70 hours of cumulative data entry time. For those costs we would estimate the librarians who input the data make \$12 an hour.

Then there would be the time it takes for ALVA to look through the data entries to determine which ones will be approved or unapproved based off the criteria (before our theorized code is implemented), which could take anywhere from a simple Google Image search to check photos of library buildings to numerous time consuming ways to verify libraries. We estimate maybe one person could approve one library location within 15-20 minutes, then for the hundreds of locations resulting it maybe 80-100 hours. For ALVA employees or workers, we estimate they make \$15 an hour with 4 people working on this portion. With our trust verification credits we estimate to cut those hours down to 40, as there will be less time needed to check accuracy.

2) Hardware and software

The labor and price of purchasing an ArcGIS subscription to create the online maps would be done by ALVA, for the project we use the UW students account.

Anyone using the system or inputting library location points would also need to have access to computers, a place that has computers, and electricity. This could be in the form of a computer or a mobile smartphone.

An Esri ArcGIS subscription can range from \$700- \$3,800 a year for GIS professional product. With ArcGIS online being considerable cheaper. The software for coding python is free, but the

time it took to create it are anywhere from 30-40 hours accumulation. For the sake of the project we will estimate that the technology needed for inputting the crowdsourced data by the librarians/users is a basic \$30 a month phone plan. Then the estimated 1 employee of ALVA who created the software and hardware update make \$20 an hour and rent technology \$50 a month. No change in cost with our project added.

3) Implementation

The cost of implementation for this project would be in the process of the labor it takes to do the project, how ALVA will use it and then used by AfLIA. There could be costs such as compensation for employees of either organization. The implementation cost would be like the costs mentioned in the Database section of this above. But only looking at ALVA's working compensation. For ALVA employees or workers, we estimate they make \$15 an hour with perhaps 100 hours of operation with three people working on it. With our trust credit system, it should only take 40 hours per ALVA employee to implement instead of 100.

3.2.2. Operating Costs

1) Personnel

Personnel labor cost includes the compensation of any people employed by either mentioned company, as well as the cost of time we committed to the project. This would only be recurring costs for the two organizations once our involvement is concluded. We estimate maybe one person could approve one library location within 15-20 minutes, then for the hundreds of locations resulting it maybe 80-100 hours. For ALVA employees or workers, we estimate they make \$15 an hour with 4 people working on this portion. With the trust credit system verifying library locations we estimate it should only take 1 person about 50 hours.

2) Overhead

General upkeep to the organization, such as ArcGIS program/software subscription for both ALVA and TASCHA and Esri ArcGIS subscription can range from \$700- \$3,800 a year for GIS professional product. Our project will not change this cost.

3) Maintenance fees

The maintaining of the program we've designed should be minimal, keeping our code simple will help that. People of the organization will need to maintain the maps and tables to reflect future data collected. We estimate maybe 4 hour a year to upkeep our trust credit system. For ALVA employees or workers, we estimate they make \$15 an hour.

3.3. Benefit Cost Analysis

Database

	Librarian/users	ALVA Labor
Time(hrs)	70	400
Money(\$)	840	6000

Database with our trust verification credits

	Librarian/users	ALVA Labor
Time (hrs)	70	160
Money (\$)	840	2400

Hardware

	ArcGIS Subscription	ALVA Labor	Alva Tech Rent	Smart phone plan
Time (hrs)	0	40	0	0
Money (\$)	3800	800	30	50

Implementation

	ALVA labor
Time (hrs)	300
Money (\$)	4500

Implementation with our trust verification credits

	ALVA labor
Time (hrs)	120
Money (\$)	1800

Personnel

	ALVA labor
Time (hrs)	400
Money (\$)	6000

Personnel with our trust verification credits

	ALVA labor
Time (hrs)	200
Money (\$)	3000

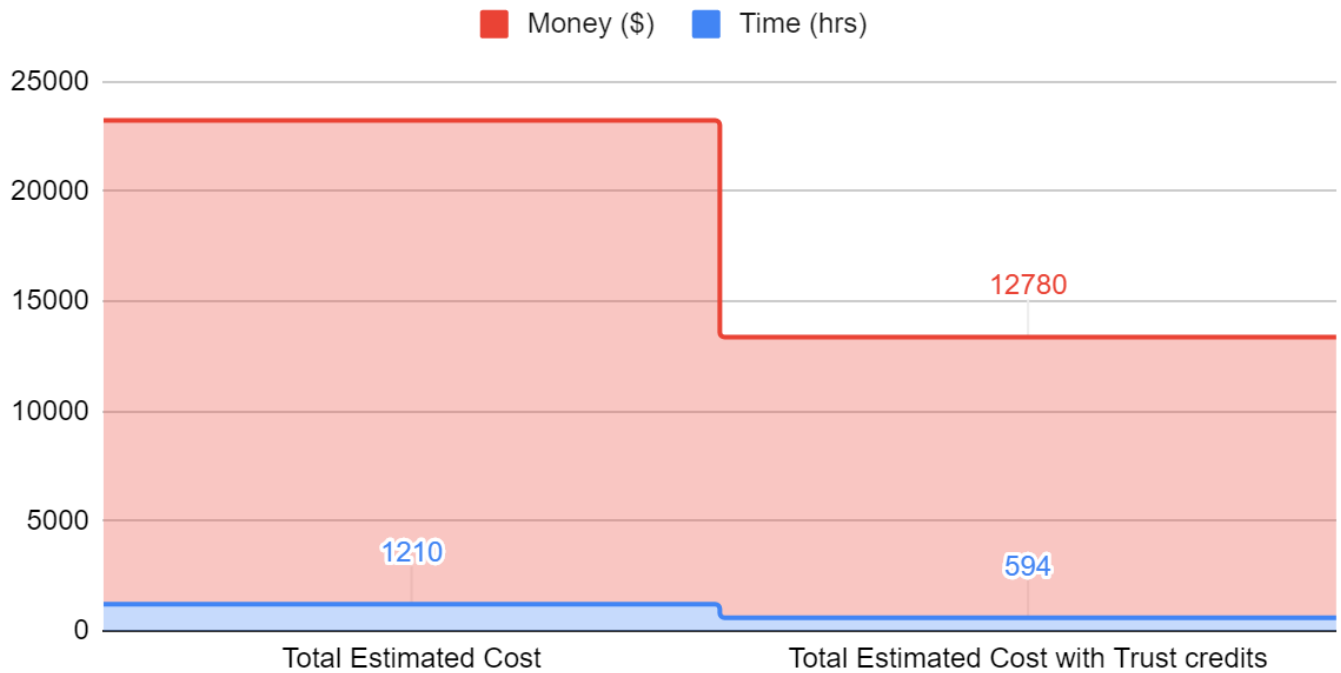
Maintenance with our trust verification credits for a year

	ALVA labor
Time (hrs)	4
Money (\$)	60

Total

	Total Estimated Cost	Total Estimated Cost with trust verification credits
Time (hrs)	1210	594
Money (\$)	22020	12780

Total Estimated Cost and Total Estimated Cost with Trust credits



4. Data Development

Data development was all on crowdsource, meaning data development was dependent on librarians/users opening the online map and entering their library locations. The issues that were the most prominent in the project is in the data quality issues. When we first looked at the data, there were many locations, or data points, that had been entered that were either duplicate entries, or had already been deemed unreliable or rejected from ALVA, or had been entered and were awaiting verification to ensure the data was reliable or pending. That's where our trust credit (verification credits) implementation would come in, to provide a sense of trust to users in their decision making about which data to might be more accurate.

One of the reasons we determined data quality might be an issue, was in the process of inputting a new location. The process of entering a new point, includes picking a point on the map then answering a series of questions. It was determined that this series of questioning along with the sequence was slightly misleading. It was also determined that the table schema was also in need of restructuring as many of the fields were unused or many of the entries may need to be drop down instead of an open input type.

For a definite project plan set up, we theorized the data table to be restructured, and the trust verification credits framework was designed in a way that was fast and direct to ensure an efficient user input experience. Then we designed a rough program that would allow users/libraries to log in to the system to input their points as well as to verify other locations to gain credits. We decided on a credit system to determine a user and library location verification value, that is then used to determine how verifiable the user or library can be deemed.

5. Workflow Implementation

Since the project done by ALVA was already initially designed and created, our directive was to improve a part of the data quality assurance process. Creating the system of trust verification credits, the theorized alteration of the data table and a rough program example for a way to log into the system was our major role.

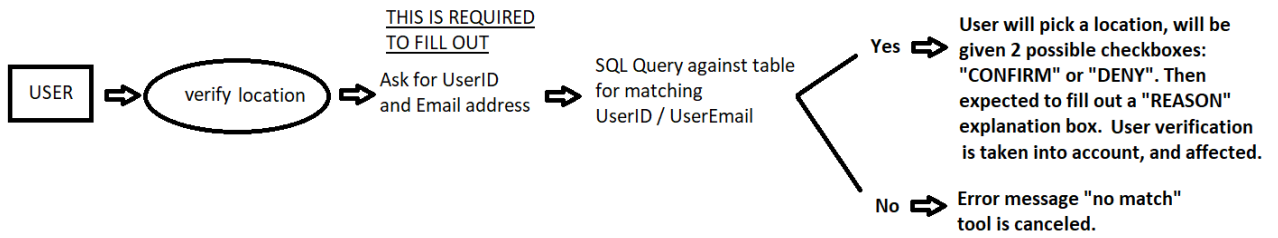
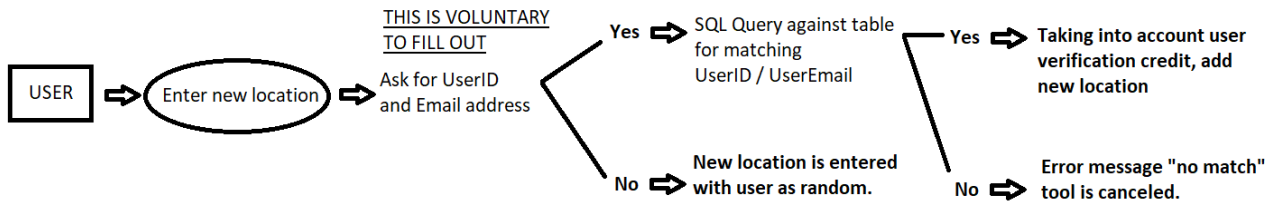
In the beginning, the table was slightly ambiguous, as there were many fields that seemed to be a duplicate to another very similar field, or there seemed to be several fields that didn't look as if they were being used. Looking at the data that had been collected, there were many library locations that seemed to be obvious duplicates, as they were the same name, same address input, but different location on the map. Some of these points had been verified while many of them had been rejected, while some were still pending for verification. While it made sense to keep the rejected points shown on the map, but in a different color, it was slightly misleading visually.

We spent a lot of time researching and questioning our community partner contact regarding what was expected and how the program was working and why aspects were designed the way they currently were. Then after spending some time examining the type of data we were using, it was decided the best avenue given time and resources was peer data verification, thus creating our trust credit system (verification) idea.

After working through a few different ways to go about the trust credit idea, the best was theorized to be something that differentiated between different users, with who was inputting data. As well as fixing the data table, having a login system would help the implementation of the trust credit system. The table had to be restructured to add the credit system, the user interface was theorized for both entering new library locations but also for the librarian/user to verify other locations as well.

Our conceptual idea for this design was to end it with a mock implementation into ArcGIS desktop, due to time restraints we were only able to create a rough idea. With our trust credit framework, to be designed as a theorized idea that ALVA can use to evolve their data accuracy program.

A concept was built for the program and a trial edition via python for an ArcGIS desktop version. Due to the changes caused by the only way to log into ArcGIS online (AGOL) being either through ESRI or through ALVA, it was determined that the login policy had to be ignored. Several blogs were read and inspected for a work around, but another common line of arguments seem to indicate that Python programming was for desktop version and for AGOL you had to utilize JavaScript. Given the time we had for this project the only choice we had was to theoretically build this script in whatever language it needed to be written in, considering our login restrictions. Our model can be seen below:



6. Results

The goal of this project was to find a way to help ALVA look at the sub-Saharan Africa crowdsourced data and decide whether or not to trust the information based off our verification process. Based on this statement we had met this theory goal. We created a solid conceptual plan that met the goal for ensuring data trustworthiness for a crowdsourced collection platform and created a rough plan for how a login system could be implemented.

As we looked at what was produced from the data and how it was working up to this point, there were library locations that to this current development, had been verified. However, there were points that needed to be restructured due to misleading information. Based on our project it was determined the best design was to utilize the library/users as if they were industry co-workers and hoped there would be a professional pride involved in ensuring data accuracy.

We hit times where we found it to be challenging, including our lack of experience with programming. It seemed that the more blogs that were read an equal number of contradictions were found. Early on we wanted to produce a usable program, but due to time constraints and the inability to determine programming issues, we had to settle for a conceptual product. Once we found our focus we moved forward and we seemed to figure out how to get through problems and issues efficiently and quickly find our next point of focus.

6.1. Data table Alteration

For our trust credit system to be implemented, we recommend that the data table be altered to fit the system. Keeping in mind this is an evolving system and would need updates over time to match the changes in data gathered. Once the table is edited it will help in the approval process that ALVA does for library locations. Thus, gathering the information needed for the trust verification credits (verification) to run on. Our alteration can be seen below:

PhotosAndFiles	hyperlink
Comments	text lib type
VerifyStatus	%
LibraryAddress	text
LibraryName	text
Date	date
UserStatus	%
UserEmail	text
UserName	text
UserPassword	text
UserID	text
ContactEmail	text
LibraryContact	text
Community	text
TypeOfLibrary	Drop down
Country	text

6.2. Trust Verification Credits

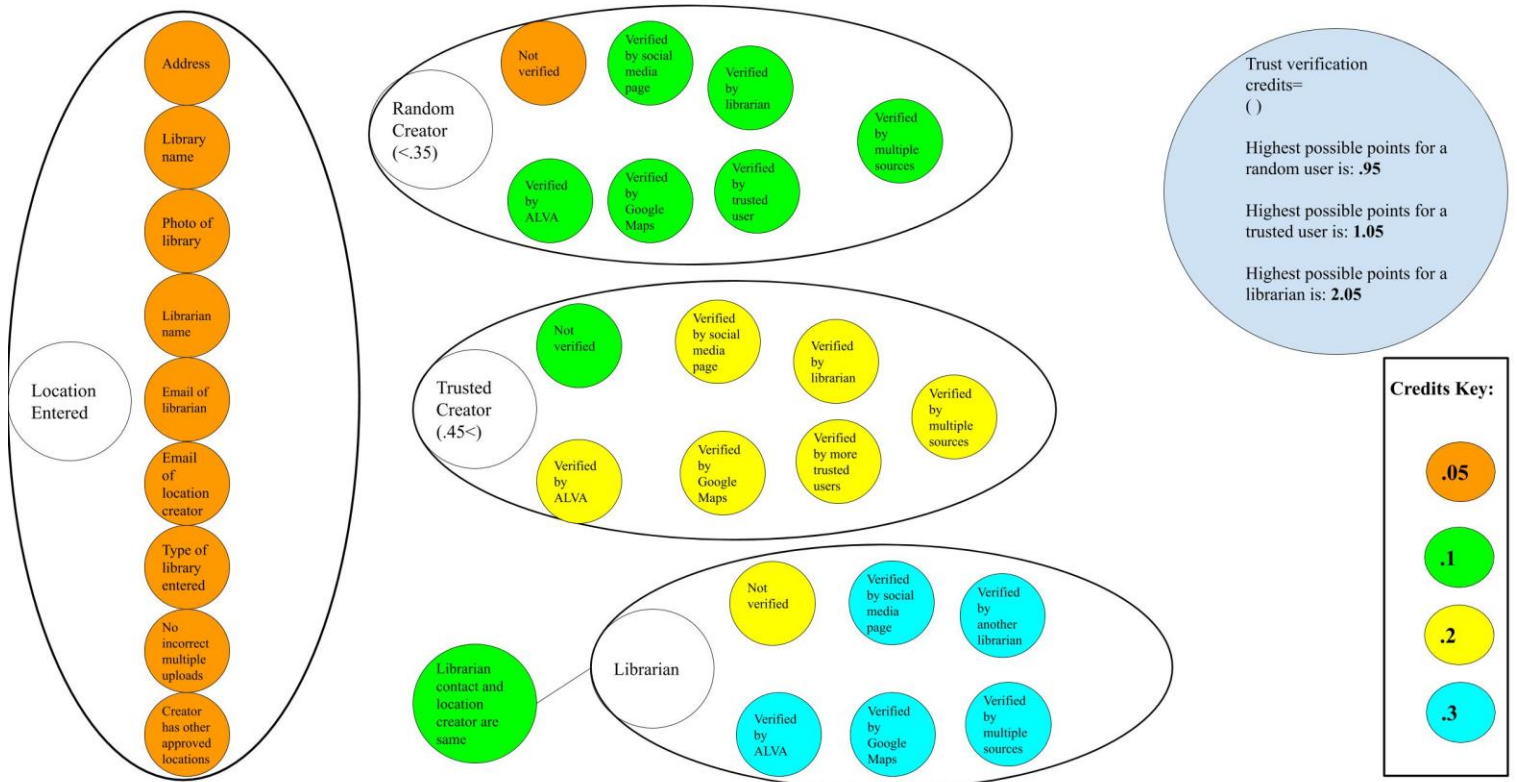
Designing the system of trust verification credits was a necessity for data to be seen as more or less trustworthy. This will help make a more quantifiable way to measure how trustworthy a library location is.

Trust Credits	Description
.05	Location has complete address
.05	Location has complete library name
.05	Location has photo posted in the contribution
.05	Location has library contact name
.05	Location has email address of library contact
.05	Location has contact of who created point
.1	Location has librarian and "point created by" information are the same (name and email)
.05	Location has stated type of library
.05	The same location hasn't been uploaded more than once
.05	User has contributed more than one approved location
.05	Contributed by a random user, and not verified by any other source
.1	Contributed by a random user, is verified by ALVA
.1	Contributed by a random user, is verified by a social media page
.1	Contributed by a random user, is verified by Google Maps from their imagine posted
.1	Contributed by a random user, is verified by a librarian

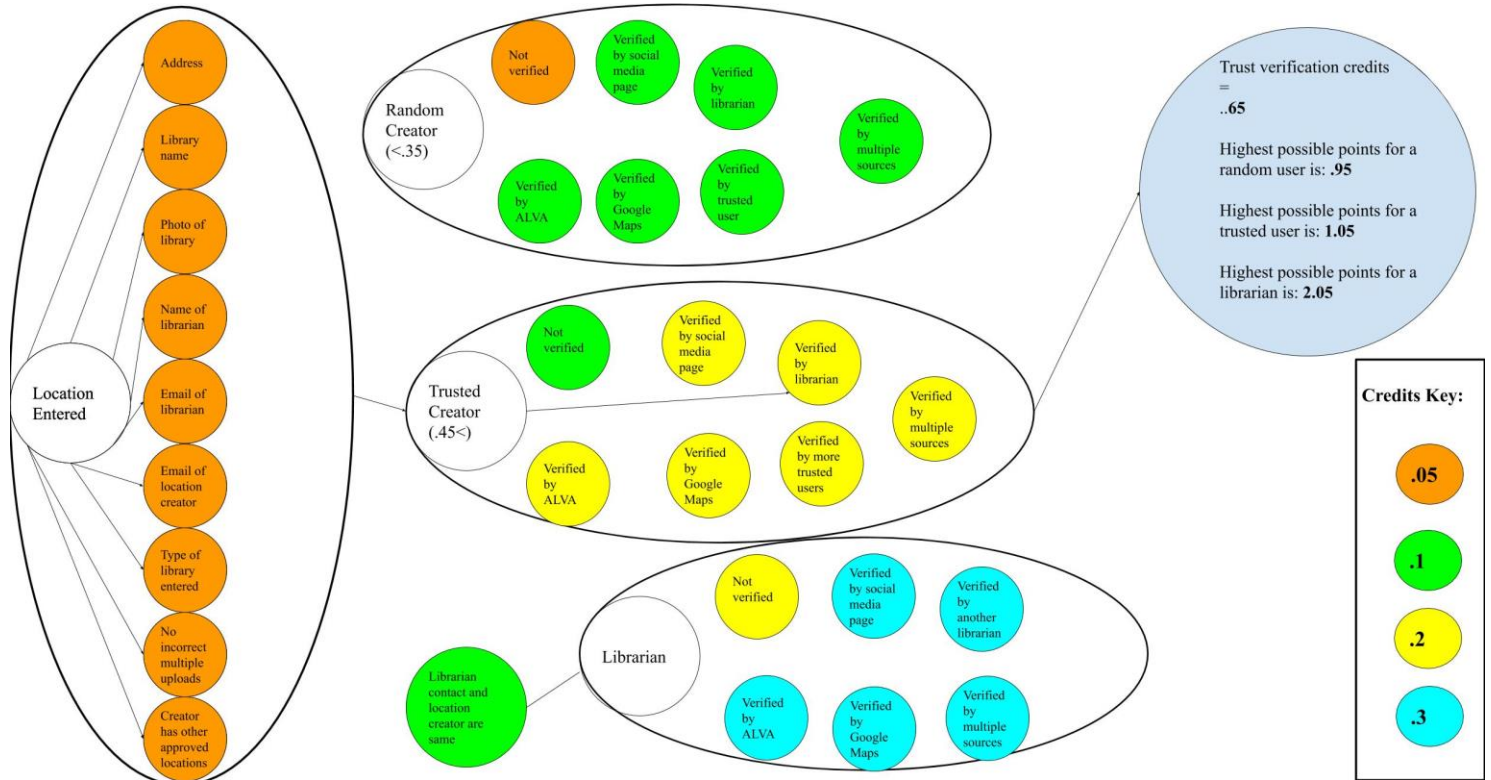
.1	Contributed by a random user, is verified by a trusted user
.1	Contributed by a random user, is verified by more than one source
.1	Contributed by a trusted user, and not verified by any other source
.2	Contributed by a trusted user, is verified by ALVA
.2	Contributed by a trusted user, is verified by a social media page
.2	Contributed by a trusted user, is verified by Google Maps from their imagine posted
.2	Contributed by a trusted user, is verified by another trusted user
.2	Contributed by a trusted user, is verified by a librarian
.2	Contributed by a trusted user, is verified by more than one source
.2	Contributed by a librarian, and not verified by any other source
.3	Contributed by a librarian, is verified by ALVA
.3	Contributed by a librarian, is verified by a social media page
.3	Contributed by a librarian, is verified by Google Maps from their imagine posted
.3	Contributed by a librarian, is verified by another librarian
.3	Contributed by a librarian, is verified by more than one source

6.2.1. Trust Verification Credits Framework

The framework for the trust credit (verification) system is a comprehensive visual of how we theorize the trust verification credits to work. We hope that it will streamline the process of evaluation for ALVA. Keeping in mind that this is an ever-evolving system and changes may be made accordingly to the data gathered. In addition, it is important to remember that this is in no way a formal decision on which data to trust, but to help users make an informed decision based off information gathered.



An example for a library location entered into the system may look like this:



6.3. Login Program

It was initially thought that logging in was going to be required to find a way to verify and ensure data accuracy through the concept of crowdsourced data collection. The concept was designed that by entering a new location a user was suggested to login but would not be required. To ensure that users were not pushed away from using the application, the only time required for an actual login would be for a user to verify a location. Jason advised against a login process as he was afraid it would drive away potential users, but we could not see around forcing a login without creating a way to account for data accuracy with data inputs.

Being that this map was an AGOL application it seemed the best way to test was to download a desktop version then when a working prototype was created, show it to Jason Young for a review and acceptance before modifying the online version. A working desktop version was never completely created for lack of programming background. However, after reading different blogs online at ESRI's blogging community online, it was determined that users had to login to either ESRI or the organization to perform any kind of edit on data.

It was obvious that we had to find a 'work-around' from logging into the map, after reading more blogs we became aware of the idea of checking unique attributes. After some more research and some deliberating it was determined the best way to work around the login was to check UserID against user Email. UserID may not be unique, but email addresses must be unique and for a user to be able to match a UserID to a User Email is as close to logging in without creating a potentially unpopular login process.

7. Conclusions and Recommendations

For this project our main goal was to determine a way that the crowdsourced data gathered by ALVA could be verified as accurate. Our system of building a sense of trust about the data could possibly help with that. Based on what we knew about the project up to this point, it seemed the map was being used by possibly over 65 different individuals. Out of 835 location inputs, there were 64 inputs that did not add a UserID in the table and 763 were inputted by one individual. Since we started working on this project there were approximately 59 new locations added to the map. We hope that designing a more in-depth system of verification will help ALVA in the long term of their project.

7.1. Conclusions

We believe, based on client requirements, that the final plan for peer data verification could potentially be a stable operation for verifying crowdsourced data. Aside from the one user with the UserID, many users do not have a UserID. This indicates that future users may not appreciate signing up for a membership as it was indicated by our client, due to the hassle that making login accounts often create. Besides that, our design choice to require users who want to confirm locations, to log in as a requirement would be greatly beneficial to the locations data by creating even more validity.

One area where more research could have been done to improve or changed the plan, would be in better understanding of programming and tools. There were a few key points where the project path had to veer in order to create a proper deliverable.

7.2. Recommendations

In the end we believe that ALVA would find our project a good launching point to aid in a more comprehensive verification process over time. With crowdsourced data, being that it is using the help of so many people it would only make sense to make the verification process also done by all those people. We recommend understanding that this is an evolving project and shouldn't be taken "as is". Our hope is that ALVA can examine our trials and errors and use our project as a reference for validating their crowdsourced data in the future.

8. References

Advancing Library Visibility in Africa (ALVA) information

Their ArcGIS online map

<https://uw.maps.arcgis.com/apps/CrowdsourceReporter/index.html?appid=8a4e7f85628f405c81d8cbbcb6fee110>

Their project website:

<https://osf.io/685hm/>

Technology and Social Change group (TASCHA) website about ALVA:

<https://tascha.uw.edu/publications/preliminary-activities-of-the-advancing-library-visibility-in-africa-project/>

African Library & Information Associations & Institutions (AFLIA)

<https://web.aflia.net/>

Antenucci, J. C. et al. 1991. "Chapter 4: Benefits and Costs." *GIS: Guide to the Technology* , pp 66- 72

5 types of benefits:

Type 1: Quantifiable efficiencies in current practice, or improvements to existing practices such as cost reduction of labor.

Type 2: Quantifiable expanded or added capabilities to be more effective with work activities.

Type 3: Quantifiable, unpredictable events that can take advantage of geographic information.

Type 4: Intangible benefits related to intangible advantages.

Type 5: Quantifiable sale of information.

Nkem Osuigwe, Gertrude Mulindwa. 2018. "African public libraries driving development: AFLIA Leadership Academy projects." *Libraries as Centers of Community Engagements for Development*.

<http://library.ifla.org/2316/1/s01-2018-osuigwe-en.pdf>.

9. Technical Appendices

The following is an outline of people, sources not categorized as references, and programs and software used.

Title page picture is from ALVA's ArcGIS online map found in section 8.

9.1. Appendix A: Contacts

Community Partner:

Jason Young- from Advancing Library Visibility in Africa (ALVA)

youngjc2@uw.edu

Authors:

Amy Beller- MGIS Student

amybeller@outlook.com

Mark Johnson-MGIS Student

markfjohnson81@outlook.com

9.2. Appendix B: Blogs

Geonet ESRI Community used for many references

<https://community.esri.com/thread/>..... (depending on blog title)

9.3. Appendix C: Programs and Software

-ArcGIS Online

<https://www.esri.com/en-us/arcgis/products/arcgis-online/overview>

-ArcGIS Pro

Version 2.3., Copyright 2018 Esri Inc. All rights reserved

-ArcGIS API for Python - used as reference

<https://developers.arcgis.com/python/>

-Python

<https://www.python.org/downloads/>