The main job of a reference librarian is to know where to go for information resources needed by their clients and many of these resources are housed in databases. Fortunately there are programs offering developing countries free or reduced cost access to their journals and books. However librarians in developing countries struggle with finding these organizations. The overarching problem is that they have to deal with a huge or total lack of Internet connection, and this is one of the biggest enablers to the digital divide that is currently and rapidly growing between first and developing countries. This project aims to connect journal providers who provide their content at free or reduced cost to underserved countries to librarians in these countries by means of a hybrid database called Journal Access for Underserved Areas JAFUA.

Headings:

Database design – Database development

Scholarly periodicals – Scholarly publishing

Acquisition of electronic journals – Electronic journals

Access to information – Information needs

Developing countries – International cooperation
JOURNAL ACCESS FOR UNDERSERVED AREAS (JAFUA): DESIGNING A DATABASE FOR UNDERSERVED COUNTRIES FOR THE WIDERNET PROJECT

By Laura Ashcraft

A Master's paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Information Science.

Chapel Hill, North Carolina
July, 2015

Approved by:

___________________________
Cliff Missen, M.A.
Advisor
Table of Contents

List of Figures ................................................................................................................................. 2

Introduction/background .................................................................................................................. 3
  High level problem .......................................................................................................................... 3
  Statement of need ............................................................................................................................ 4
  Resolve the problem ......................................................................................................................... 5

State objectives .................................................................................................................................. 8

Literature review ............................................................................................................................... 9

Project Scope .................................................................................................................................... 21

User Analysis .................................................................................................................................... 23
  Core group ...................................................................................................................................... 23
  Secondary group ............................................................................................................................. 25
  Potential future groups ................................................................................................................... 27

System Analysis .................................................................................................................................. 28

Description of original data: turning data into information .................................................................. 28

Design database .................................................................................................................................. 31
  Conceptual modelling ...................................................................................................................... 31
  Logical modelling ............................................................................................................................. 32
  Physical modelling ............................................................................................................................ 33

Using Workbench to create tables ...................................................................................................... 37

Exploring the JAFUA .......................................................................................................................... 38

Management & Security ..................................................................................................................... 41

Problems Encountered ....................................................................................................................... 42

Future Plans ......................................................................................................................................... 46

References .......................................................................................................................................... 49

Appendices .......................................................................................................................................... 53
  Appendix 1  Project team and roles ................................................................................................. 53
  Appendix 2  Project Resources .......................................................................................................... 54
  Appendix 3  Developing Countries Initiatives & Open Access Collections ...................................... 56
  Appendix 4  Data Definitions (each Table) ....................................................................................... 57
List of Figures

Figure 1: Sub-Sahara's Internet usage (per 100 inhabitants) in 1999............................................. 15
Figure 2: Sub-Sahara's Internet usage (per 100 inhabitants) in 2010............................................. 15
Figure 3: Challenges of teaching and learning ICT courses in HLIs............................................. 18
Figure 4: Initial conceptual model for project ............................................................................... 33
Figure 5: Initial logical model for project...................................................................................... 34
Figure 6: Sample of m:m relationship tables ................................................................................. 35
Figure 7: Physical modeling done with Workbench........................................................................ 38
Introduction/background

This project aims to connect journal providers who provide their content at free or reduced cost to underserved countries with librarians in these countries by means of a hybrid database called Journal Access for Underserved Areas JAFUA.

High level problem

Programs have been created for the specific purpose of providing high quality reference resources to underserved populations. The problem is that librarians in these location have to navigate through myriad barriers to get to these programs. The main barriers this project focuses on are described below:

Connecting libraries in underserved areas with programs: The main job of a reference librarian is to know where to go for information resources needed by their clients, and many of these resources are housed in on-line databases. While there are many programs that can provide relevant information many librarians remain unaware of them. Also the content providing programs may not have the adequate means of advertising their services to these locations (due to lack of ICT and telecommunication infrastructure).

Unreliable Internet access: Most of these underserved communities have to deal with a huge or total lack of Internet connection. And this is one of the biggest enablers to the information and digital divide that is currently and rapidly growing between first and developing countries.
**Funding:** On-line DBs and journals are expensive. For example, in 2014 the University of Manchester, UK, paid a total of £1,543,204 for a year's subscription of Elsevier [25]. Fortunately there are programs offering developing countries free or reduced cost access to their journals and books. However librarians in these countries struggle with accessing these programs, again due to a lack of awareness of these programs.

**ICT capacity:** Fostering awareness and understanding the complex options and requirements for these programs requires high-level librarian, technology, legal, and management skills.

**Statement of need**

**What is the problem?** A librarian gets many information requests for research, and the most necessary component in research is accessing journals. While some patrons may know the specific journal they need, the majority do not and seek out a librarian to aid in finding the best resources. For most countries this would not be a problem, the librarian would simply fire up the internet, search through databases, and then acquire the relevant materials. However, for African librarians in developing countries, this is not so simple a process. According the Internet World Stats, Africa makes up 9.8% of the world's Internet users [41]. This is due to the high costs in connecting to the net (via satellite, Wi-Fi, or broadband). For these librarians, this is a huge problem.

Individual programs or journals that provide subsidized access have a matrix of qualifications to meet and requisite infrastructure. Librarians are required to assess and apply for each. A user-centric finding tool, that presents only options available to them given their circumstances, does not yet exist.
In order to meet the information needs of their patrons, a database housing free and subsidized journal information would be helpful. It is the goal of this project to create such a database.

**Resolve the problem**

**High level solution:** The WiderNet Project decided to take on this challenge and develop a database that would house the manifold collections that are either free or are discounted through programs between journal publishers and certain institutions that meet criteria favorable to developing countries. This database, it was imagined, would give the African librarian the information they need to access these repositories, be it the link to an Open Access collection or to organizations that are actively involved in providing these countries with access to high quality content via registration through their institutions.

**The WiderNet Project:** The WiderNet Project seeks to address critical information issues within developing countries. An example is their flagship project: the eGranary Digital Library which is a database with over thirty million documents and acts as an intranet, thus getting surmounting Internet connectivity issues.

**My involvement:** I worked as a digital librarian volunteer for this project during my first semester at the School of Library and Information Science at the University of North Carolina, Chapel Hill. Before this experience I had known intellectually about what is called the digital divide, but I had never experienced it. I fell in love with the mission and knew that I wanted to do my master's project in this field. The project director told me he needed help with this project.
The WiderNet Project librarians had created a spreadsheet of data that needed converted into a database. This data consisted of information about programs aimed at giving certain countries journal collections free or at a greatly reduced cost. These programs partner with top publishers (i.e. JSTOR, Wiley, Elsevier) to produce journal packages for countries that match the criteria of being a developing country. Each program had its own criteria or restrictions. Besides these larger, more formal programs, the WiderNet librarians had also combed other journal collections that met such criteria as: legitimacy (the content provider provided actual and trustworthy material. Websites were checked against the open access predatory list: Scholarly Open Access: http://scholarlyoa.com/publishers/), subjects were relevant (information was relevant to the user's needs), and availability (the content could be accessed for free or at a reduced subscription).

The next step in this mission was to take this data and turn it into a database that can be accessed both on and offline.

**Products created:** Two products will be created for this project: a database and an interface via a web site.

The **database** will track journal organizations and open access repositories identified and assessed by the WiderNet Project digital librarians along with users and experts worldwide. The database will store information to make it possible for users (librarians in underserved areas) to search for journals either individually or through collections based on that librarian's location and capacity.

The **interface** will allow librarians to access this information through a web site especially made for this project that will be accessible both on-line and off-line. The
website will have step by step instructions on how to navigate and interact with the database.

My master's project will consist of identifying the requirements and developing a database prototype, while the rest of the team (team members and roles are explained in the Appendix) will create the interface.
State objectives

The objectives for developing the Journal Access for Underserved Areas (JAFUA) will be handled in five stages:

1. Clean the data
2. Design the database based on the data
3. Develop the physical database
4. Create views (read in Future Plans)
5. Test database by generating MySQL statements (read in Future Plans)
Literature review

In order to create a database that will deliver up to date information about journal programs specifically tailored for developing countries, I had to answer three questions:

1. Is there an information and digital divide in Africa, specifically sub-Saharan Africa?
   And if there is, then why?
2. Is there an Internet access problem in Sub-Saharan Africa? If so then why?
3. What other databases have been built to address this problem?

These questions helped develop the criteria I used for my literature review. The criteria in brief:

- Papers describing the information and digital divide between underdeveloped countries and developed countries.
- Papers describing Internet access and reliability issues in these countries.
- Papers describing database projects addressing these countries.

1. Is there an information and digital divide in Africa, specifically sub-Saharan Africa?
   And if there is, then why?

   It has been well documented that there is an information gap between rich countries and poor countries. First a review of the problem of developed countries with regard to the acquisition of publications. This will help highlight the gap in developing countries.

   **Information divide:** The problem with journal acquisitions and libraries. The rising cost of journal subscriptions by university libraries to top publishers has become
what is known as the "serial crisis". These costs have become so high that top universities in America have started considering submitting their research to open access publishing. An example of this can be seen in a communiqué written from Harvard's Faculty Advisory Council to "Faculty Members in all Schools, Faculties, and Units". This letter stated in no uncertain terms that acquiring materials from top publishers "have made the scholarly communication environment fiscally unsustainable and academically restrictive"[35]. They even made a statement/plea to "move prestige to open access" [35]. A white paper written by the University of North Carolina Chapel Hill provided research given by the Association of Research Librarians that looked at subscription costs given from their member libraries. Between 1986 and 2003, “the price per subscription of serials rose by 215%” [37]. Another study was done in the UK that looked at the rise of subscription costs, these researchers asked over 100 universities to provide their annual subscription payments to top publishers within the last 5 years (2010-2014). Publishers represented were: Elsevier, Wiley, Springer, Taylor & Francis, Sage, Oxford University Press, Cambridge University Press, Nature Publishing Group, Royal Society of Chemistry, and the Institute of Physics Publishing. The results showed that the prices per publisher and between schools were uneven. For example: University of Manchester paid a total of £7,395,406 for Elsevier within these 5 years, whereas the University for the Creative Arts spent £8,273 [25]. Even though the data doesn't explicitly explain which collections these universities bought, a Google search on each university tells us what they teach and a possible explanation for the contrasting prices could lie in the fact that the University of Manchester is a STM (Science, Technology, and Math) as well as offering other research areas, whereas the University For the Creative Arts is not a STM
university. The UNC white paper explained that “STM journals show some of the steepest prices and price increases" [25].

So if the top countries in the world are struggling with subscription costs, it is no wonder that developing countries in Africa are at such an information disadvantage. But these issues are nothing new for developing countries in Africa, for there is a long history of and well documented literature on the information poor and information rich. Some of the barriers towards accessing and producing information lie is educational and training inefficiencies [33], institutional financial lack [6], governments political interests in suppressing information (Burkett, 2000), lack of physical infrastructure (roads, telecommunication lines) [18], and not least international politics (IMF pressuring governments to privatize infrastructure, visa requirements that restrict African scholars from participating at international conferences, as well as trade agreements such as WIPO and TRIPS that want to restrict fair use within electronic information) [22]. All of these factors (and many more) are present and worst, exacerbated when it comes to discussing the digital divide. Here is an example of how unequal ICT access and use is in developed countries. Muir and Oppenheim researched factors that inhibit "universal" access within developed countries:

- telecommunication infrastructures need to be completed;
- people with disabilities need technological assistance to overcome barriers of a standard computer interface;
- users’ set-up and access costs;
- lack of physical access;
- a perceived lack of relevant content; security concerns;
- lack of skills and training; and
• illiteracy (both the inability to read, and the inability to use IT properly).

[26]

Again, if these are factors in developed countries, then underdeveloped countries have yet another barrier to overcome. To understand the digital divide in developing countries in Africa, a better understanding of the Internet and ICT use and access within underdeveloped Africa is needed.

2. Is there an Internet access problem in Sub-Sahara Africa? If so then why?

Some might think that lack of ICT and Internet capability could be resolved in having the hardware, such as a computer. Yet this is an erroneous assumption for "computer technology represents only the pinnacle of a whole array of prerequisites" [7]. Requisite technology comes in the form telecommunication infrastructure, funding, political and economic support, and ICT literacy (education and training). The following section will discuss the various barriers to ICT access and implementation within developing countries in Africa.

**Sub-Sahara access to Internet data:** Analyzing the data provided by the World Bank, we can see that in 1999, top Countries in North America and those in Europe and Central Asia had a significant advantage in internet use over those countries in Sub-Sahara Africa:
Figure 1. Sub-Sahara's Internet usage (per 100 inhabitants) in 1999. Compare this number to those of the developed nations in North America and Europe and Central Asia. From World Bank data, mapped with Google's Public Data Explorer, retrieved from http://tinyurl.com/nfvzsth. Last updated Jun 9, 2015.

And it isn't until 2010 that Sub-Saharan Africa's internet usage gets above 10%:

Figure 2. Sub-Sahara's Internet usage (per 100 inhabitants) in 2010. Compare this number to those of the developed nations in North America and Europe and Central Asia. From World Bank
The Unreliable Internet

Internet cost: SSA countries, due to the lack of internet subscription by population (not enough people to drive the cost down as well as not enough Internet company competition), must pay far more for internet service. In a study that explored medical professionals’ use of ICT with SSA it was remarked that "Internet service can cost around $50 per month, while the annual salary is often less than $1,000" [11].

IEEE Spectrum looked at the University of Kinshasa, Democratic Republic of Congo, and saw costs fifty times that of developed countries: "While a university in Germany might pay about US $4,000 per month for 1 gigabit per second of bandwidth, a school in Kenya can expect to pay $200,000 for the same service.” [9].

Electricity: Many Sub-Saharan countries struggle to provide consistent electricity to a fraction of their population. Even the larger, more prosperous countries, like Nigeria and South Africa, experience frequent and significant outages. Reuters Africa reported South Africa's main electrical supplier, Eskom, had to initiate rolling blackouts due to their poorly maintained electrical infrastructure. Very recently this company had to implement a last effort "load shed" (where power is deliberately cut off from users in order to stave off total collapse within the power grid) and this is said to be costing the SA economy " $1.7 billion (R20 billion) and $6.8 billion (R80.1 billion) a month" [30].

Fiber-optics: An example of vulnerable data infrastructure lies in the submarine cables enabling much of coastal Africa broadband connectivity. Whether by earthquakes or accident [8], ships [17], or by malicious people [27], these cables, if damaged, can take out broadband service in multiple countries. Wired magazine reported on this
vulnerability describing the fallout of a cut cable: "A cut cable off the coast of Alexandria in 2008 left Egypt, India, Pakistan and Kuwait in the dark. A 2006 earthquake in Taiwan damaged several cables and cut off communication to Hong Kong, South East Asia and China" [8].

**Slow Speeds:** Researchers studied traffic patterns within two South African rural areas (Macha, Zambia and Dwesa, South Africa) and found the following factors that contributed to poor network and internet connectivity:

- VSAT internet connectivity only produced "a committed download speed of 128 kbps bursting to 1 Mbps and a committed upload speed of 64 kbps bursting to 256 kbps "[23].
  - Compare this number, 128kbps, to the average American download speed of 18,000kbps (home) and 5,530 kbps (cell phone) [34].
- Very regular power failures per month "lasting an hour to a few days".
- Large traffic load goes to web applications such as Facebook, Skype and web browsing [23].
- Malware attacks due to out of data virus signatures, lack of technicians to maintain the computers as well as a faulty perception of just whose responsibility it is to secure personal computers [23].

**SSA poor ICT infrastructure:** Authors Rumanyika and Mashenene analyzed 13 papers describing the challenges in utilizing ICT within Tanzania HLIs (higher learning institutions) and noted the critical challenges had in common. They used Unified Theory of Acceptance and Use of Technology (UTAUT) to test whether ICT was being positively or negatively utilized by collecting the factors that were obstructing ICT
utilization. Their initial review formed a hypothesis that these challenges would make UTAUT aspects (performance expectancy, effort expectancy, social influence, and facilitating conditions) negative and therefore overall ICT implementation (use, learning and teaching) would be negatively affected. Their findings show the critical factors that negatively impact ICT utilization with Tanzanian HLIs [33]:

![Figure 3. Challenges of teaching and learning ICT courses in HLIs.](image)

This figure illustrates the critical factors inhibiting ICT implementation. Adapted from "Challenges for Teaching and Learning Information and Communication Technology Courses in Higher Learning Institutions in Tanzania: A Review" by J. D. Rumanyika and R. M. Galan, 2015, Information and Knowledge Management, 5, p. 8, Copyright 2015 by IISTE.

Here the most critical factor inhibiting ICT implementation is poor infrastructure. This entails poor bandwidth capability and an overall lack of appropriate telecommunication structures as the biggest barrier [33]. A quick look at the World Bank data shows that the percentage of Tanzania that had access to electricity, as of 2012, was 15.3%. Compare this number to the US 100% [38]. The most recent data for electric
consumption (kWh per capita) shows that in 2011 the score was 92kWh. Compare this number in the same year to the US: 13,240 kWh [39].

Within figure 3, two other factors stand out: ICT access and ICT competency. Factors contributing to the lack of ICT hardware are funds, lack of adequate technical experts, lack of bandwidth, and poor supporting telecommunication infrastructure [33]. As for the ICT competencies, Ruminyaki and Galan conclude that there must be a "poor instructors-students ratio in HLIs in Tanzania" [33].

**Conclusion about digital divide:** Recapping the ICT and Internet problem in developing countries of Africa: Internet access is either too expensive, ICT may not be supported by necessary telecommunication infrastructure, and training and maintaining technical experts can be problematic. Besides being slow, the Internet that relies on electricity and fiber optic cables is vulnerable to electrical blackouts and torn underwater cables. And finally not least is the issue of political support in governments not interested in an informed populace via ICT [3] [4] [10].

3. What other databases have been built to address this problem?

Getting information to rural areas in the world has been a challenge to which many researchers and organizations have devoted themselves. In 1990 SatelLife delivered the first peer-reviewed article discussing the use of Vitamin A for measles via the satellite HealthNet in select African countries [32]. Other organizations have begun developing countries' initiatives with the goal of getting up to date journal information to these countries. Below is a small sample of such initiatives. A much more thorough list will be discussed in Exploring the JAFUA section as this database project will be cataloging them.
Online digital libraries and databases:

- **Edinburgh University Press (EUP)**
  - Publishes books and journals. Through the Developing Countries Initiatives certain countries can either gain free access to their products or get a reduced subscription. EUP also partners with the Programme for the Enrichment of Research Information (PERii), which offers certain African developing countries a discounted subscription price to their journals [13].

- **Research4Life**
  - This is an overarching program that hosts 4 programs (HINARI, AGORA, OARE, and ARDI) aimed at providing free or discounted subscriptions to eligible developing countries. HINARI (Access to Research in Health Programme) provides health journals, AGORA (Access to Global Online Research in Agriculture) provides agricultural journals, OARE (Online Access to Research in the Environment) provides environmental journals, and ARDI (Access to Research for Development and Innovation) provides research and development journals. [2].

- **JSTOR**
  - Through their Developing Nations Access Initiative JSTOR provides eligible developing countries free or reduced cost to their journals [14].

- **OXFORD**
  - Their developing countries initiative provides journals to certain developing countries at a free or reduced cost. They also partner with INASP
(International Network for the Availability of Scientific Publications), Research4Life, and EIFL (Electronic Information for Libraries) [12].

**Hybrid and offline digital libraries and databases:** While access to the Internet is accessible yet unreliable (for now), alternatives have been made such as digital libraries via hybrid technology (online and offline), and CD-ROM and offline databases. As there are many initiatives in Africa focusing on this problem, this paper will use one country's program initiatives as an example.

Below is a list of digital library projects in Ghana:

- **George Padmore Research Library on African Affairs**
  - Place and Purpose: Accra. In 2007 digital preservation and digitization efforts focused on the historical files about Kwame Nkruma's presidency [36].

- **University of Cape Coast Digital Library (UCC)**
  - Place and purpose: University of Cape Coast. In 2005 the university library began digitizing the university's educational materials to aid in distance learning [36].

- **Gramophone Records Museum and Research Centre (GRMRC)**
  - Place and Purpose: Centre for National Culture in Cape Coast. In 2002 the museum received funds to digitize Ghanaian music from 1920' on [36].

- **ITC Digital Library**
  - Place and purpose: Kwame Nkrumah University of Science and Technology in Kumasi, Ghana. Holds geo-information (GIS) and remote sensing information (RS) gathered between the Kwame Nkrumah University of
Science and Technology (KNUST) and International Institute for Geo-Information Science and Earth Observation in the Netherlands [24].

- eGranary Digital Library
  - Place and purpose: Ashaiman, Ghana (year unknown). Augments primary and secondary education [28].

- Librii
  - Place and purpose: Accra. A new KickStart project aimed at providing a public library via an "e-hub made from a modified shipping container that contains high-speed computers along with other digital tools and … a public plaza equipped with Wi-Fi" [40]. This project is still awaiting some final funding.

In conclusion, while there are online and offline products that provide developing countries access to journal information, there is not one that is focused on indexing all journal databases aimed at these countries that can be received free or at a reduced cost either through non-profit or for profit organizations. It is the goal of this project to begin such a database. Further discussion of how this project will accomplish will be discussed in the Exploring the JAFUA section.

**Research conclusion:** Due to lack of funding, poor or vulnerable ICT infrastructure and capacity underserved areas need an alternative means of connecting to high quality journals. Hybrid databases can provide such a means in that they contain valid and relevant content accessed both online and offline.
Project Scope

This is a holistic description of the data evaluation, database critical components, creation process, and lack/future components. More detailed descriptions will be discussed in the system analysis section.

Data (Evaluate the given data, getting the data ready for a DB): The data in the spreadsheet must be cleaned: atomizing the data which consists of standardizing the data. After that is normalizing the data and then from there the entities and their attributes can be developed.

Design (Design (ideation) and conceptualization, database table creation): This stage consists of modeling the entities to understand their relationships. Conceptual models are the rough sketches, the logical models are a more detailed sketch including the entity's attributes, and the physical model contains the most detail housing the data types.

Project components (What this project entails: DB prototype and logical design): Three products will be created for this project: a relational database and two interfaces via a web site. My master's project will be the production of a prototype relational database.

Future components: What this project lacks: Interfaces (will be created later by the WiderNet Project). The interfaces will allow librarians to access this information and the second interface will allow for the back end team to curate the collection. The
websites will have step-by-step instructions on how to navigate and interact with the database. The interfaces will be designed in tandem with my project by the team.
User Analysis

Core group

Brief description of the African librarians: The African librarian has the same education and tasks that their counterparts all over the world does. They are graduates from an accredited university and are tasked with helping their patrons (reference), organizing and curating the collections (subject specialists), administration (acquisitions, managerial, etc.), cataloging, and other tasks. The poorer and more rural the library, the more multitasking the librarian must be. This project aims to aid the librarian that needs journal information for her or his country, and by her or his institution (mainly a university).

Snapshot of an African librarian (Persona):

- Knowledge and experience
  - College education, domain of knowledge wide
- Computer/IT experience or knowledge
  - Medium computer knowledge and competencies. Main experience is data processing (using documentation tools such as the Office suite). Very high experience with mobile use and technology. Has low-medium knowledge of ICT hardware installation and maintenance. If there is a breakdown in the ICT hardware and/or software, she may try to troubleshoot for simple solutions or problems she is familiar with. But she is comfortable with calling in IT.
- Level of experience with the task
High level of experience in utilizing databases for information gathering purposes. However, due to Internet connectivity issues and low browser rendering capability, she may be used to simpler databases rather than the complex ones.

Psychological characteristics

- Highly empathetic, great listening skills, observant, high communication skills.

Task Analysis: What they will need to do to use the system

The librarian will need to be able to have a medium competency with ICT, specifically utilizing a computer that is connected to the JAFUA. They will need to be familiar at least on an experiential level the 3-tier architecture: the browser connects to a databases via a server. Even though this database can be accessed offline, there will be an online site as well that will be more up to date.

User scenario:

A new librarian in a university in Rwanda is given the task of cataloging free health journal content that her college can apply for. The college specifically wants access to three journals: Oxford, NEJM (New England Journal of Medicine) and HINARI (Access to Research in Health Programme), but she is to also find other journals with the same high caliber of these collections. So she uses the JAFUA. Once she gives her location (Rwanda) she can give the type of access she requires (free or paid), in this case she clicks free. She is given a list of collections with these criteria. She scans for the three journals she wants and sees that they are included. Clicking on those shows her that Oxford, NEJM, and HINARI all have programs that provide her country with journals for
free. The database includes links to the application pages for all three. She peruses the other collections in the list and notes the ones of interest that meet her criteria. Some of the collections have certain restrictions such as rolling-walls (new content can be accessed after a certain time period), some are in certain languages, and some may be viewed in certain formats that she may have to upload in order to view.

**Secondary group**

**A brief description:** Besides the librarian in a university, this system may be produced for primary schools up to technical colleges. Also African LIS students may employ the JAFUA either in training or in their internships.

Another group that will definitely be interacting with this JAFUA will be the back-end volunteers or employees charged with maintaining the DB and interfaces. These people may be part of the WiderNet Project both in Africa and America. Within this organization will be digital librarians whose role is to curate the entire collection, update changes to countries status via a change within a program, and add new collections and programs.

**Snapshot of a back-end developer:**

This example will outline a general back-end developer whose task is to maintain the JAFUA.

- **Knowledge and experience**
  - College student, domain of knowledge specific. Work experience part-time work as an IT team member. School projects hone his database understanding, MySQL and SQL coding as well as school training in coding in PHP and Python.
• Computer/IT experience or knowledge
  o Intermediate to high computer knowledge and competencies. Main experience with troubleshooting issues such as handled by IT departments he may have worked for. May have had undergraduate training in computer science and foundational knowledge of JAVA, C++, or some other programming language.

• Level of experience with the task
  o Low to medium experience in utilizing databases for information gathering purposes. Most of his education has been application based and projects oriented. Research was done with material either already had (books in his class) or informal Internet searches relying on Google information retrieval knowledge to get his resources. College may be the first time he has learned about designing databases and their applications.

• Psychological characteristics
  o Low to medium soft skills, including empathy. Highly skilled in linear, logical thinking: rendering communication into systematic thinking.

**Task Analysis:** What they will need to do to use the system.

The technical volunteer will need to be able to have a medium to high competency with ICT, specifically utilizing a computer that is connected to the JAFUA. They will need to be on expert level with the 3-tier architecture. As his job will be troubleshooting any issues within the database as well as interface issues, he will also need to have medium to high knowledge and skills with back-end development such as PHP, MySQL, HTML, and JavaScript.
User scenario:

An email was sent from a Rwanda librarian alerting the JAFUA team of non-working hyperlinks. The IT volunteer pulls up the appropriate mark-up page and compares the hyperlinks to those in the JAFUA tests both sets. Sure enough the links turn up a "404 Page not found" error. He goes to the original site and navigates to the desired links and finds that the links have been changed to redirect to different pages. He fixes the links in both the website markup page and the database. After he successfully tests both sets again he logs the error and emails the librarian that the links are working.

Potential future groups

For now this project is focusing on developing countries within Africa, but the Director of the WiderNet Project expects that this product could be employed in other developing countries such as India, the Americas (including North American poverty stricken schools), and even prisons. Other important groups will be the collaboration with users, publishers, and project managers via web interface and fora.
System Analysis

Description of original data: turning data into information

The original spreadsheet had developed four entities: Aggregators, Journals, Books, and Countries. Aggregators are collections of journals, either different journals in a collection or just one title in a collection. Journals contains single individual titles and will be considered the same as Aggregators (the two sheets were eventually merged). Books are published or produced by organizations involved in disseminating critical information (i.e. health care, technical, science, etc.). And Countries holds the data which organizations have created programs for underdeveloped countries. This project only focuses on the Aggregators sheet as this was the more challenging data set, also the procedures learned and successfully implemented at this stage would be the outline for the other entities.

The original Aggregators spreadsheet data:

- Website - Name of the collection
- Link - URL link to the collection
- About - A brief description of what the collection is about
- Host Index - If the collection resided in the eGranary, then this field would provide the index number to it
- eGranay - A binary field of whether the collection was hosted in the eGranary or not
- Available offline? - Describes the formats the content could be saved and then later read offline
- Open online? - A binary field of whether the collection could be read online or not
- Paid online? Discount? - This field describes which journals in the collection were free or at what cost
- Other - Some notes of interest on a collection

From this point I followed a set of procedures was developed to clean scrub the data set and which would then help me create the database. Here are the procedures:

- Correctly define the fields
  - Standardize the data
  - Atomize and normalize fields
  - Add or remove fields

Correctly define the fields: Remote collaboration and personal meetings were had with the team’s librarian in order to understand and refine data. Collaboration was done through remote software tools such as Google Hangouts, HipChat (both video and texting mediums) as well as Google Drive for document review.

Standardize the data: The spreadsheet had taken the WiderNet librarian months to compile. As such, when listing data for a field that were having some duplicated information, she would eventually start to standardize the data. The problem was that she had other tasks to attend to and so this spreadsheet would sit for days or weeks, and when she would get back to compiling data for the fields, she would sometimes set a new standard for the data. Also, as she got to know the data better, she would change the standards accordingly. Standardization came in understanding which data was truly unique and which data was a representation.

An example of standardizing: Using the field name "Available Offline?" This field has been converted to two fields to describe 1) whether or not the journal can be
accessed offline and 2) what format can be stored. But the original field had this non-standardized data:

- "Available journals have individual article links to PDF download."
- "Availability and format depends on individual journal title, but in general looks to be PDF and HTML."
- "Titles have individual article PDF links for download."

**What is the difference between a title and a journal?** Most of the fields had to be grammatically cleaned (spelling errors, punctuation removed when unnecessary), parsed (simplifying the data and removing erroneous data: atomize), standardized (figure out which data could possibly mean the same thing, in other words figure out naming conventions), and normalized (remove data from a column if it should be its own field).

This exercise was a great lesson in learning the difference between creating a database from an information standpoint and creating a database from a data standpoint (more about this in Final Comments section).

**Atomize and normalize the fields:** This stage occurred in tandem with the previous procedure. Standardizing the data led to a better understanding between what was erroneous data and that could be taken out of the data set. After the data was cleaned it was possible to discern which fields would be converted into entities and attributes.

**Add or remove fields:** Much of this occurred during normalization, however in modelling the entities and then testing them in the DBMS with queries resulted in better understanding of the entity relationships to their attributes: sets of entities were modelled in UML mockups, tested in the Workbench DBMS, and then find out which entities needed further modification (more or less attributes) or remove them entirely.
Design database

Conceptual modelling: After the spreadsheet data was cleaned, fields were converted into entities. After several versions (modifying the spreadsheet) the entities were tested for child-parent relationships as well as which would become lookup tables. Also this stage helped in figuring out the types of relationships between entities. A good resource I had for spelling out relationships can be had for free in iTunesU, by Professor Jay Jarmin of the University of Tennessee [20].

Here is an initial concept design. The finished model is different from this but this design provided the core understanding:

![Figure 4. Initial conceptual model for project.](image)

**Figure 4.** Initial conceptual model for project.
Logical modelling: Create relationships: Logical modeling helps in taking the abstract ideas of entities and formally organizing them in relationships. This model depicts primitive relationships in that there are many many-to-many (M:M) tables. However this gives me the starting point in understanding which entities will need bridge tables.

Sample (Explanation of full relationships is in the Appendix 4.): This sample describes the relationship between Collections and Journals.

In English: One collection can have many journals, and one journal can be accessed in more than one collection.

Sample relationship:

<table>
<thead>
<tr>
<th>Entity</th>
<th>Relationship</th>
<th>Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection12</td>
<td>Contains</td>
<td>Journal 29</td>
</tr>
<tr>
<td>Journal 29</td>
<td>Is accessed through</td>
<td>Collection7</td>
</tr>
</tbody>
</table>

Final analysis of relationship: this is a m:m relationship. The final tables will need a bridge table to convert the m:m to two 1:m relationships.
Physical modelling: Create parent and child tables: This is where the schema and tables as well as data types inside the DBMS were created. Listed below are the entities in this section but for a full explanation, the entities and their relationships can be viewed in the Appendix 4.

Parent tables:

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collections</td>
<td>Collections are the groups of journals. Collections can be a group by a single publisher or an aggregate of publishers.</td>
</tr>
<tr>
<td>Journals</td>
<td>Journals are the single titles of a collection. This table will be the largest of the main parent tables.</td>
</tr>
<tr>
<td>Organizations</td>
<td>Lookup table for Programs. Only some programs are run by an official organizations.</td>
</tr>
<tr>
<td>Programs</td>
<td>Programs table describe the program that negotiates with countries for certain collections of journals. This table has a 1:M relationship with Organizations table (1 Org. can participate in several Programs.) and has a M:M relationship with Countries table.</td>
</tr>
</tbody>
</table>

Table explanation: We want the African Librarian to be able to access journal information from all levels, be it accessing a single title to accessing the overall collection information, and even getting collections from Organization and Programs. So if a librarian wants to access all journals under the Organization AGORA (a collections of agricultural journals) they could query for all collections under the organization named

```sql
CREATE TABLE Collections (CollectionID, CollectionName, CollectionAbout, URL, Online, Restrictions);

CREATE TABLE Journals (JournalID, JournalName, JournalURL, JournalAbout);

CREATE TABLE Organizations (OrganizationID, OrganizationName, OrganizationAbout);

CREATE TABLE Programs (ProgramID, ProgramName, ProgramAbout);

CREATE TABLE has (CollectionID, JournalID);
```

**Figure 6.** Sample of m:m relationship tables.
AGORA. Or if the librarian wanted to look for just the title "BioChem", they could also query for that title in the database.

**Child tables:**

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>AccessTypes</td>
<td>Lookup table: This table describes how an institution can gain access to the collection.</td>
</tr>
<tr>
<td>Coverage</td>
<td>Lookup table: This table describes the text that is available: abstracts, full-text, or both.</td>
</tr>
<tr>
<td>Countries</td>
<td>Lookup table: These are the countries that our database holds program information for.</td>
</tr>
<tr>
<td>Focus</td>
<td>Lookup table: These are the subjects that a journal or collection covers.</td>
</tr>
<tr>
<td>Formats</td>
<td>Lookup table: This table describes how the journal can be viewed (HTML or PDF or both).</td>
</tr>
<tr>
<td>Languages</td>
<td>Lookup table: This table describes the languages that the journal can be viewed in.</td>
</tr>
<tr>
<td>Offlinetypes</td>
<td>Lookup table: This table describes how the journal can be viewed offline.</td>
</tr>
<tr>
<td>Publishers</td>
<td>This table describes the Collections' Publishers. (This may be a lookup table if there are duplicate Publishers for multiple Collections).</td>
</tr>
<tr>
<td>Restrictions</td>
<td>Lookup table: This table describes how a collection may be restricted from an institution. This table differs from the AccessTypes in that it tells how a collection is closed from the user whereas the AccessTypes tells how a collection can be accessed. For instance, a collection may be accessed for free with registration but only to certain Vocations.</td>
</tr>
<tr>
<td>Roles</td>
<td>Lookup table: Roles describe the publishing type category that the collection is being displayed under. Publisher both owns the collection and grants permission to its contents. A Broker is a third party that has bought from another publisher collection(s) and has the right to grant users permission access. A Guide neither owns nor has the right to grant access, they simply hold a list of collections as a reference guide and can point to the collection's location for users to access. For Guides, material ranges from free access to subscription.</td>
</tr>
</tbody>
</table>
Table explanation: Tables AccessTypes, Focus, Formats, Languages, OfflineTypes, Restrictions and Roles are children tables to the Collections table. These describe how to access a collection and how that collection can be read either online or offline and in what language. Roles tells the user whether or not the collection is being represented by a publisher within the database or is the source for the collection acting as an approved gateway (approved being that it has purchased the content of the collection to provide to users) or if the collection is simple a list of links that point to the journals for the user to further research. Focus is the subject of the Collection. (In the future some of these table relationships will be changed. More on this in the Problems Encountered section.)

Tables Coverage and Publishers belong to the Journals table. And Countries belongs to the Programs table.

**Bridge tables (M:M converting to 1:M relationships):**

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>collections_has_accesstypes</td>
<td>This table describes which Collection have which types of AccessTypes.</td>
</tr>
<tr>
<td>collections_has_focus</td>
<td>This table describes which Collection have which Foci.</td>
</tr>
<tr>
<td>collections_has_formats</td>
<td>This table describes which Collection have which type of Formats.</td>
</tr>
<tr>
<td>collections_has_journals</td>
<td>This table describes which Journals belongs to which Collections.</td>
</tr>
<tr>
<td>collections_has_languages</td>
<td>This table describes which collection have which Languages.</td>
</tr>
<tr>
<td>collections_has_offlinetypes</td>
<td>This table describes which Collection have which types of OfflineTypes.</td>
</tr>
<tr>
<td>collections_has_programs</td>
<td>This table describes which Collections are being hosted by which Program.</td>
</tr>
<tr>
<td>collections_has_restrictions</td>
<td>This table describes which Collections have</td>
</tr>
</tbody>
</table>
which type of Restrictions.

collections_has_roles | This table describes which Collections have which type of Role.

Journals_has_coverage | This table describes which type of Coverage belongs to which Collection.

Journals_has_publishers | This table describes which Publisher belongs to which Collection.

programs_has_countries | This table describes which Countries has access to which programs.

Table explanation: These tables represent the relationships between the parent and child tables.

Create relationships: This is the current model. There are additional tables as well as the insertion of bridge tables.

Figure 7. Physical modeling done with Workbench.

Create Lookup tables: Lookup tables acts as a data dictionary that holds a strict set of data. The purpose of this is to constrain the possible inputs (choices) within a table.
The lookup tables are located on the far right in the previous image. Relationships are drawn for a high level understanding but are not necessary. A full description of tables can be viewed in the Appendix 4.

Using Workbench to create tables

Schemas and tables were created using the Workbench GUI rather than MySQL code due to future considerations. The majority of employees/volunteers that will maintain this DBMS will be digital librarians rather than IT personnel. Therefore it was felt that the user-friendly aspects of the Workbench program would be a better development tool.
Exploring the JAFUA

As was stated in the Introduction, the goal of this database is to provide developing countries in Africa a means of tracking journal accessibility via programs directed towards giving the free or reduced cost access to their collections, or tracking collections that are freely available. This database will also track restrictions of these collections as pertains to that librarian’s country. Below is a fuller description of these programs and collections.

Resources available:

• Criteria for journal database entry into the JAFUA:
  ○ All resources have at least some free journal content available.
  ○ Journals are either peer-reviewed or vetted by the WiderNet Project’s digital librarians.
  ○ Resources can cover a wide range of topics, not exclusive to Africa.

• Content providers:
  ○ There are three types of providers, which are called Roles in the JAFUA:

  • Publishers: organizations that publish and produce the content.
  • Brokers: does not publish the content but has the right to produce it, right is given by the publishers. Some aggregators are collections that are publishing their own content and has the right to provide other publishers’ works.
  • Guides: neither publishes or has the right to provide the content. These are usually websites that act as a reference guide to resources of interest and provide absolute links to these resources.
• Information about access:
  ○ Online content can be accessed via an Internet connection.
  ○ Offline content can be accessed either through the eGranary (if the eGranary contains it) or can be downloaded in various formats from the Internet to access offline. Also some content can be emailed or sent to the librarian via mail orders. Information on how to access the content will be included in the content description within the DB.
  ○ Information about providers such as URLs, organizations involved in the content's production, formats and languages the content can be read in, and any type of restrictions placed on the collection (i.e. rolling-walls, restricted by country, registration access, or no restrictions in the case of Open Access publishers) will be provided for each collection in the DB.
  ○ The countries listed in the JAFUA are African developing countries as listed by the organizations that have developing country initiatives. The DB will keep track of changes in countries' status.

Below are an example list of major programs providing developing countries free or reduced cost content that is included in the JAFUA. To see the details of each program a fuller list is given in Appendix 3.

• Developing countries programs:
  ○ Edinburgh University Press (EUP)
  ○ Research4Life
  ○ JSTOR Developing nations access initiative
  ○ OXFORD Developing Countries Initiative
- IMF (International Monetary Fund) Free Subscription Policy
- NEJM (New England Journal of Medicine) Free Access from Outside the U.S.
- PTOLEMY
- RSC (Royal Society of Chemistry) Archive for Developing Countries

- Open Access publishing:
  - Directory of Open Access Journals (DOAJ)
  - APS
  - eJDS (electronic Journals Delivery Service)
Management & Security

Maintenance will be done by the WiderNet Project administrator in charge of this site. A team of digital librarians both locally and remotely will coordinate for data curation while an IT employee or volunteer will look after the security of the database and site. The team will need to ensure that the online site will have proper security such as separate server-side scripting locations, appropriate permissions per roles, updating security software, using secure FTP for file upload and download, and adequate backups. The website developers will make sure that any sensitive data (such as database login and password information) in the PHP files will be properly hidden. Also when designing the forms input data will need to be carefully constrained so as to not let erroneous data corrupt the database.
Problems Encountered

**Initial data:** As was stated, the initial data set received was a non-standardized, non-atomized spreadsheet of data. As I began this project after my first class of databases, this came as a shock. Before this experience I had only created databases from higher level information standpoint (more on this in Final Comments). Here I had to reverse engineer a database from the data out. I had to really understand each field and the cells beneath them in order to know which data could be standardized and which data was truly unique.

**Collaborative issues:** The main tools used with the content specialist (the project's digital librarian) was an internet application called HipChat. This application allowed for instant chatting as well as video conferences. This was a great tool but was not the best when it came to collaborating on the spreadsheet. For this task Google Drive was used. Google Drive also allowed document storage via Google's cloud under a person's account. But this had a couple of problems. The first being, at the time, there was a 400,000 cell limit. This may seem like it should be enough but in collaborating for edits, this was not enough. However Google has recently updated the Drive's cell capacity to 2 million cells. Understanding the strengths and weaknesses of collaborative tools will be useful for future collaborations with project team members and users.

**DB needs a new perspective (from Access to Country):** When building the database, I was looking at a lot of access type information (how free was a collection), and this formed my database creation with the perspective of access. Meaning that I
focused on how to access different collections and how free or with what type of restriction would hinder that access. Later I had a meeting with the director, Professor Missen (also my master's project advisor), and he told me that the perspective needs to be from the Countries perspective. He also noted how three parent tables would have access and restriction information that could also differ. So instead of having the AccessTypes, Restrictions, and Countries be child tables to Collection only, I have to reconfigure the relationships so that Collections, Programs, and Journals have this relationship. This perspective will also need a scoring mechanism attached to the AccessTypes table.

The finished database should be able to get the appropriate programs and collections from a given country and then automatically apply a scoring output so that the user will see a list of collections with the "free-est" score at the top of the list to the most restricted collection showing at the bottom. These calculations will need to be created by the PHP/interface team. This new perspective will mean that I go back to the database physical models and re-arrange the countries table to reflect this new perspective.

These new relationships will need to be attached to the bridge tables:

**programs_has_countries**
- CountryID
- ProgramID
- AccessID
- RestrictionID

**collections_has_programs**
- CollectionID
- ProgramID
- AccessID
- RestrictionID
- CountryID
collections_has_journals

- CollectionID
- JournalID
- AccessID
- RestrictionID
- CountryID

**Need automated programs for capturing Collection data:**

Focus: This is just too much work for a (future) content analyst to have to deal with. We need to find a way to have a program automatically read from a website and capture key words that it knows to be a focus subject. This focus list will need to be created, a good resource for this will be the Library of Congress. This table will be deleted and shelved for future purposes.

General about-ness data: The same program can be applied here when ingesting collections information from a website.

**Reconfigure relationships:**

- Formats, OfflineTypes, and Languages will need to be related to the Journals table rather than the Collections table. The reason for this is that a journal has the more specific of these fields and can easily be indirectly queried from the Collections table.

- Restrictions and AccessTypes
  
    - Restrictions and AccessTypes: These two tables reflected similar information and need to be better differentiated. A new understanding of the tables read:
      
      - AccessTypes describe how to get access: Free, Open Access, Free w/ registration, Restrictions Apply, Reduced Cost, Full Subscription. And when there is a restriction, then the Restriction table will be activated.
• Restrictions are barriers to access: describe what these restrictions (if any) are: None (for Free or Open Access), Institution, Vocation, etc.

**Need a letter to Publishers for requesting metadata of their journals:** Only a small portion of publishers had their collections organized in a csv file. A letter asking publishers for their metadata will be extremely useful as this will cut down a huge time in data processing upon the future digital librarians that will be curating and adding to this database.
Future Plans

Entities and their relationships have been created. Sample SQL statements have been successfully tested. The future tasks for the JAFUA are the following:

Find out which collections/journals are in the eGranary: Some collections may reside in the eGranary, the WiderNet Project's digital librarians and volunteers will find which are indexed and relay that information to the JAFUA. This will be useful both end users who have an eGranay, that way they can simply access the material.

Upload the rest of the journals (+100,000): A small sample of the data was uploaded for modeling and database testing purposes. The journal table is the most complex one with regards to relationships, so after further modifications are done to the existing model, testing for this table can begin. Below are what still needs to be incorporated into the JAFUA:

- Upload the Books data
- Upload the Countries data
- Upload the Journals data (this is a separate set of journal data, individual titles only)

Implement the 3-tier architecture (Client-Server-Browser): My part of the project is to develop the database, the Web team will take care of the interface (web-site, PHP server connection and protection).

Product implementation: Once the database and interface are ready it will be ready for user testing. The users will consist of volunteers from the WiderNet Project as well as African librarians.
Establishing a connection to underserved country initiative programs: Plans for the future entail our team establishing a communication with these programs as well as others. For the myriad of reasons this paper has already discussed, users that most need these types of programs may not be able to navigate to these sites when they are able to browse the Internet. This DB can act as an advertisement for content providers to get the word out to their intended users. Also content providers will be given forms and understand what information the JAFUA team needs to better display their collection information (these needs consist of metadata that organizations have created to maintain their own journal collection databases).

Updating the JAFUA database: Once the databases has been successfully deployed updates will be done via asynchronous methods (Web interfaces will be used to collect updates directly from partners and publishers).
**Project Conclusion**

Data vs Information: In my database classes, I was usually given a high level idea from which to create a database. I call this creating a database from an information standpoint because data has not yet been gathered, and all tables (entities and attributes) must be conceptually created with all possible data perspectives in mind. For example, if my teacher gave me the task of creating a database for a future shoe store, then I would have model/guess what might my entities be. And then I would input dummy data to test the logical models and reconfigure as needed. The important thing to remember in this scenario is that I don't have any real-world data to grow my database around. All ideation, conceptual and physical modelling must be done from a high-level, outside-in standpoint (information).

On the other hand, this project gave me very solid data. Instead of having a clean slate for modelling first and then inputting data to test the models, I had to create my database from the inside-out so to speak. I had to first understand the data, which was not easy to do at first for there was a lot of standardization that needed to happen. Normalization had to occur after the data was cleaned and standardized. And after that then could I start physical modelling with data types. Relationship testing happened once I could input some test data into the Workbench, and this usually concluded more normalization within the spreadsheet. This was a more iterative process than my experience with creating a database from an information stand-point, and thus a very valuable exercise.
References


Appendices

Appendix 1  Project team and roles

Cliff Missen – Project Director for the WiderNet Project at the University of Chapel Hill, NC
Laura Ashcraft – Database developer (prototype)
Billy Gagon – (Future) IT troubleshooter
Joan Ferguson – Digital librarian
Connor Michos – (Future) PHP programmer
Appendix 2  Project Resources

Software
- MySQL Workbench
  - This is the DBMS my project will be using to create the database
  - http://dev.mysql.com/downloads/workbench/5.2.html
- Microsoft Excel
  - This is the spreadsheet tool my project uses to process the raw data for the project
- Microsoft Word
  - This is for writing formal reports such as my project reflections and final paper
- Notepad++
  - This is a tool for modifying data
  - http://notepad-plus-plus.org/
- Microsoft OneNote
  - My main tool for note taking for this project. Allows for organizing data, thoughts, workflows, and connecting with team members
- OmniGraffle
  - Mac tool: for database conceptual and logical drawing
  - https://www.omnigroup.com/omnigraffle
- HipChat
  - This is a collaborative application, helps keep me in touch with the project members

Literature
- MySQL 5.7 Reference Manual
  - A reference manual for using Workbench
- Database Systems
  - Book: Fundamentals of Database Systems by Ramez Elmasri
  - For understanding basic to advanced database concepts
- MariaDB
  - A reference for understanding the MariaDB
- CSCI 2020 Database Fundamentals
  - Online videos: Dr. Jay Jarman (iTunesU)
  - For understanding basic to advanced database concepts

Tools
- MySQL For Excel
An add-in for Excel that allows me to modify my database tables from the Workbench within an Excel spreadsheet

**Computer Languages**
- MySQL
  - A necessary language for constructing and interacting with the data in my database
  - [http://www.w3schools.com/sql/](http://www.w3schools.com/sql/)
- PHP
  - A necessary language for getting data in and out of my database
  - [http://www.w3schools.com/PHP/](http://www.w3schools.com/PHP/)

**Server**
- UNC Server
  - Registered student and staff access
  - For connecting my database to a server
Appendix 3 Developing Countries Initiatives & Open Access Collections

Developing countries programs:

- Edinburgh University Press (EUP)
  - Publishes books and journals. Through the Developing Countries Initiatives certain countries can either gain free access to their products. EUP also partners with the Programme for the Enrichment of Research Information (PERii) which offers certain African developing countries a discounted subscription price to their journals [13].

- Research4Life
  - This is an overarching program that hosts 4 programs (HINARI, AGORA, OARE, and ARDI) aimed at providing free or discounted subscriptions to eligible developing countries. HINARI (Access to Research in Health Programme) provides health journals, AGORA (Access to Global Online Research in Agriculture) provides agricultural journals [2].

- JSTOR
  - Through their Developing Nations Access Initiative JSTOR provides eligible developing countries free or reduced cost to their journals [14].

- OXFORD
  - Also has a developing countries initiative that provides their journals to certain developing countries at a free or reduced cost. They also partner with INASP (International Network for the Availability of Scientific Publications), Research4Life, and EIFL (Electronic Information for Libraries) [12].

- IMF
  - The International Monetary Fund has a free subscription policy for eligible developing countries to their information resources [19].

- NEJM
  - New England Journal of Medicine gives free access to their journals to eligible developing countries [1].

- PTOLEMY
  - The Ptolemy Project provides health information to Africa to eligible developing countries [28].

- RSC
  - The Royal Society of Chemistry provides access to their journal archives to eligible developing countries [30].

Open Access publishing

- Directory of Open Access Journals (DOAJ)
  - This is an online repository of open access journals "covering all areas of science, technology, medicine, social science and humanities. [15].

- APS
  - American Physical Society provides free access to their journals focused in physics [5].

- eJDS
  - The electronic Journals Delivery Service, provided by the International Centre for Theoretical Physics provides free access to their journals in physics and mathematics via email attachments provided that the user registers (which is free) [16].
Appendix 4  Data Definitions (each Table)
This section will organize the tables as follows:

- Parent tables are non-indented
- Child tables are indented once
- Bridge tables are indented twice

(Parent table: Collections. Child tables: AccessTypes, Focus, Formats, Languages, OfflineTypes, Restrictions, and Roles)

CREATE TABLE `collections` (`CollectionID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'pk for the Collections table', `CollectionName` varchar(245) DEFAULT NULL COMMENT 'Name of the collection', `CollectionAbout` varchar(2000) DEFAULT NULL COMMENT 'Describes the collection, kind of journals, what kind of oaccess (free, OA, subscriptions, programs involved with, etc.)', `URL` varchar(245) DEFAULT NULL COMMENT 'URL for this collection', `Online` tinyint(1) DEFAULT '1' COMMENT 'Simple yes or no to say if the content is had online', `RestrictionInformation` varchar(2000) DEFAULT NULL COMMENT 'This gives further information to any restrictions described by the AccessType table.', PRIMARY KEY (`CollectionID`)) ENGINE=InnoDB AUTO_INCREMENT=140 DEFAULT CHARSET=utf8 COMMENT='Collections are the groups of journals. Collections can be a group by a single publisher or an aggregate of publishers.'

- CREATE TABLE `accesstypes` ( `AccessID` int(11) NOT NULL AUTO_INCREMENT, `AccessName` varchar(245) DEFAULT NULL COMMENT '1 - Free with registration: content is free if the user registration is free.', PRIMARY KEY (`AccessID`) ) ENGINE=InnoDB AUTO_INCREMENT=30 DEFAULT CHARSET=utf8 COMMENT='Lookup table: This table describes how an institution can gain access to the collection.'

  o CREATE TABLE `collections_has_accesstypes` ( `CollectionID` int(11) NOT NULL, `AccessID` int(11) NOT NULL, PRIMARY KEY (`CollectionID`, `AccessID`), KEY `fk_accesstypes_has_collections_collections1_idx` (`CollectionID`), KEY `fk_accesstypes_has_collections_accesstypes1_idx` (`AccessID`), CONSTRAINT `fk_accesstypes_has_collections_accesstypes1` FOREIGN KEY (`AccessID`) REFERENCES `accesstypes` (`AccessID`) ON DELETE NO ACTION ON UPDATE NO ACTION, CONSTRAINT `fk_accesstypes_has_collections_collections1` FOREIGN KEY (`CollectionID`) REFERENCES `collections` (`CollectionID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='This table describes which collection have which types of AccessTypes'

- CREATE TABLE `focus` ( `FocusID` int(11) NOT NULL AUTO_INCREMENT, `FocusName` varchar(45) DEFAULT NULL, PRIMARY KEY (`FocusID`) ) ENGINE=InnoDB AUTO_INCREMENT=258 DEFAULT CHARSET=latin1 COMMENT='Lookup table: These are the subjects that a journal or collection covers.'
CREATE TABLE `collections_has_focus` ( `CollectionID` int(11) NOT NULL, `FocusID` int(11) NOT NULL, PRIMARY KEY (`CollectionID`, `FocusID`),
  KEY `fk_collections_has_focus_collections1_idx` (`CollectionID`),
  KEY `fk_collections_has_focus_focus1_idx` (`FocusID`),
  CONSTRAINT `fk_collections_has_focus_collections1` FOREIGN KEY (`CollectionID`) REFERENCES `collections` (`CollectionID`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_collections_has_focus_focus1` FOREIGN KEY (`FocusID`) REFERENCES `focus` (`FocusID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=latin1
COMMENT='This table describes which collection have which Foci'

CREATE TABLE `formats` ( `FormatID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'PK for the Formats table', `FormatName` varchar(200) DEFAULT NULL COMMENT 'Describes how the content is to be read (i.e.: PDF, HTML, offline, etc.).', PRIMARY KEY (`FormatID`) )
ENGINE=InnoDB AUTO_INCREMENT=47 DEFAULT CHARSET=utf8
COMMENT='Lookup table: This table describes how the journal can be viewed (HTML or PDF or both).'

CREATE TABLE `collections_has_formats` ( `CollectionID` int(11) NOT NULL, `FormatID` int(11) NOT NULL, PRIMARY KEY (`CollectionID`, `FormatID`),
  KEY `fk_formats_has_collections_collections1_idx` (`CollectionID`),
  KEY `fk_formats_has_collections_formats1_idx` (`FormatID`),
  CONSTRAINT `fk_formats_has_collections_collections1` FOREIGN KEY (`CollectionID`) REFERENCES `collections` (`CollectionID`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_formats_has_collections_formats1` FOREIGN KEY (`FormatID`) REFERENCES `formats` (`FormatID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8
COMMENT='This table describes which collection have which type of Formats'

CREATE TABLE `languages` ( `LanguageID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'PK for the Languages table', `LanguageName` varchar(145) DEFAULT NULL COMMENT 'gives a list of languages that the content can be read in', PRIMARY KEY (`LanguageID`) )
ENGINE=InnoDB AUTO_INCREMENT=398 DEFAULT CHARSET=utf8
COMMENT='Lookup table: This table describes the languages that the journal can be viewed in.'

CREATE TABLE `collections_has_languages` ( `CollectionID` int(11) NOT NULL, `LanguageID` int(11) NOT NULL, PRIMARY KEY (`CollectionID`, `LanguageID`),
  KEY `fk_collections_has_languages_languages1_idx` (`LanguageID`),
  KEY `fk_collections_has_languages_collections1_idx` (`CollectionID`),
  CONSTRAINT `fk_collections_has_languages_collections1` FOREIGN KEY (`CollectionID`) REFERENCES `collections` (`CollectionID`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_collections_has_languages_languages1` FOREIGN KEY (`LanguageID`) REFERENCES `languages` (`LanguageID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT
• CREATE TABLE `offlinetypes` ( `OfflineID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'PK for this table', `OfflineName` varchar(200) DEFAULT NULL COMMENT 'this is a list of different means of getting content to the user offline.', PRIMARY KEY (`OfflineID`) ) ENGINE=InnoDB AUTO_INCREMENT=20 DEFAULT CHARSET=utf8 COMMENT='Lookup table: This table describes how the journal can be viewed offline.'

• CREATE TABLE `collections_has_offlinetypes` ( `CollectionID` int(11) NOT NULL, `OfflineID` int(11) NOT NULL, PRIMARY KEY (`CollectionID`,`OfflineID`), KEY `fk_collections_has_offlinetypes_offlinetypes1_idx` (`OfflineID`), KEY `fk_collections_has_offlinetypes_collections1_idx` (`CollectionID`), CONSTRAINT `fk_collections_has_offlinetypes_collections1` FOREIGN KEY (`CollectionID`) REFERENCES `collections` (`CollectionID`) ON DELETE NO ACTION ON UPDATE NO ACTION, CONSTRAINT `fk_collections_has_offlinetypes_offlinetypes1` FOREIGN KEY (`OfflineID`) REFERENCES `offlinetypes` (`OfflineID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='This table describes which Collection have which types of OfflineTypes'

• CREATE TABLE `restrictions` ( `RestrictionID` int(11) NOT NULL AUTO_INCREMENT, `RestrictionName` varchar(45) DEFAULT NULL, PRIMARY KEY (`RestrictionID`) ) ENGINE=InnoDB AUTO_INCREMENT=12 DEFAULT CHARSET=latin1 COMMENT='Lookup table: This table describes how a collection may be restricted from an institution. This table differs from the Access types in that it tells how a collection is closed from the user whereas the Access types tells how a collection can be accessed. For instance, a collection may be accessed for free with registration but only to certain Vocations.'

• CREATE TABLE `collections_has_restrictions` ( `CollectionID` int(11) NOT NULL, `RestrictionID` int(11) NOT NULL, PRIMARY KEY (`CollectionID`,`RestrictionID`), KEY `fk_collections_has_restrictions_restrictions1_idx` (`RestrictionID`), KEY `fk_collections_has_restrictions_collections1_idx` (`CollectionID`), CONSTRAINT `fk_collections_has_restrictions_collections1` FOREIGN KEY (`CollectionID`) REFERENCES `collections` (`CollectionID`) ON DELETE NO ACTION ON UPDATE NO ACTION, CONSTRAINT `fk_collections_has_restrictions_restrictions1` FOREIGN KEY (`RestrictionID`) REFERENCES `restrictions` (`RestrictionID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=latin1 COMMENT='This table describes which Collections have which type of Restrictions'

• CREATE TABLE `roles` ( `RoleID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'PK for the Roles table', `Role` varchar(200) DEFAULT NULL COMMENT 'a role is how the content is produced. This table will give a list of the different types of producers. (Publishers: make and sell their content. Brokers: do not make, but have permission from
the publishers to sell their content. Guides neither produce nor do they have permission to sell, they just point to the content.

```
CREATE TABLE `collections_has_roles` ( `CollectionID` int(11) NOT NULL, `RoleID` int(11) NOT NULL, PRIMARY KEY (`CollectionID`), KEY `fk_roles_has_collections_collections1_idx` (`CollectionID`), KEY `fk_roles_has_collections_roles1_idx` (`RoleID`), CONSTRAINT `fk_roles_has_collections_collections1` FOREIGN KEY (`CollectionID`) REFERENCES `collections` (`CollectionID`) ON DELETE NO ACTION ON UPDATE NO ACTION, CONSTRAINT `fk_roles_has_collections_roles1` FOREIGN KEY (`RoleID`) REFERENCES `roles` (`RoleID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='This table describes which Collections have which type of Roles'
```

(This table bridges two parent tables: Programs and Collections)

```
CREATE TABLE `collections_has_programs` ( `CollectionID` int(11) NOT NULL, `ProgramID` int(11) NOT NULL, `AccessID` int(11) DEFAULT NULL, PRIMARY KEY (`CollectionID`), KEY `fk_collections_has_programs_programs1_idx` (`ProgramID`), KEY `fk_collections_has_programs_collections1_idx` (`CollectionID`), CONSTRAINT `fk_collections_has_programs_programs1` FOREIGN KEY (`ProgramID`) REFERENCES `programs` (`ProgramID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='This table describes which Collections are being hosted by which Program'
```

(Parent table: Programs. Child tables: Countries.)

```
CREATE TABLE `programs` ( `ProgramID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'PK of this table', `ProgramName` varchar(245) DEFAULT NULL COMMENT 'Name of the Program', `ProgramInfo` varchar(1000) DEFAULT NULL COMMENT 'Information describing how this program works in regards to accessing it's journals.', `ProgramURL` varchar(245) DEFAULT NULL COMMENT 'the URL of the program, sometimes this can be the homepage of the collection or it can be a specific page within the collection site.'
```
Depends on how detailed the site is.', 'OrganizationID' int(11) NOT NULL, PRIMARY KEY (`ProgramID`), KEY `fk_programs Organizations1_idx` ('OrganizationID'), CONSTRAINT `fk_programs Organizations1` FOREIGN KEY ('OrganizationID') REFERENCES `organizations` ('OrganizationID') ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='Programs table describe the program that negotiates with countries for certain collections of journals. This table has a 1:M relationship with Organizations table (1 Org. can participate in several Programs.) and has a M:M relationship with Countries table.'

CREATE TABLE `countries` ( `CountryID` int(11) NOT NULL AUTO_INCREMENT, `CountryName` varchar(345) DEFAULT NULL, `GDP` int(11) DEFAULT NULL, PRIMARY KEY (`CountryID`), KEY `CountryID_index` (`CountryID`) ) ENGINE=InnoDB AUTO_INCREMENT=319 DEFAULT CHARSET=utf8 COMMENT='Lookup table: These are the countries that our database holds program information for.'

CREATE TABLE `programs_has_countries` ( `CountryID` int(11) NOT NULL, `ProgramID` int(11) NOT NULL, PRIMARY KEY (`CountryID`,`ProgramID`), KEY `fk_programs_has_countries_countries1_idx` (`CountryID`), KEY `fk_programs_has_countries_programs1_idx` (`ProgramID`), CONSTRAINT `fk_programs_has_countries_countries1` FOREIGN KEY (`CountryID`) REFERENCES `countries` (`CountryID`) ON DELETE NO ACTION ON UPDATE NO ACTION, CONSTRAINT `fk_programs_has_countries_programs1` FOREIGN KEY (`ProgramID`) REFERENCES `programs` (`ProgramID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='This table describes which countries have access to which programs.'

(Parent table: Journals. Child tables: Coverage and Publishers.)

CREATE TABLE `journals` ( `JournalID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'PK for the Journals table', `JournalName` varchar(145) DEFAULT NULL COMMENT 'Name of the journal', `JournalURL` varchar(245) DEFAULT NULL COMMENT 'gives the url for this journal', `JournalAbout` varchar(2000) DEFAULT NULL COMMENT 'Describes the journal, what kind of access (free, OA, subscriptions, programs involved with, etc.).', PRIMARY KEY (`JournalID`)) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='Journals are the single titles of a collection. This table will be the largest of the main parent tables.'

CREATE TABLE `coverage` ( `CoverageID` int(11) NOT NULL AUTO_INCREMENT, `CoverageName` varchar(45) DEFAULT NULL, PRIMARY KEY (`CoverageID`) ) ENGINE=InnoDB AUTO_INCREMENT=4 DEFAULT CHARSET=latin1 COMMENT='This table describes the text that is available: abstracts, full-text, or both.'

CREATE TABLE `journals_has_coverage` ( `JournalID` int(11) NOT NULL, `CoverageID` int(11) NOT NULL, PRIMARY KEY (`JournalID`, `CoverageID`), KEY `CoverageID_idx` (`CoverageID`) ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='This table describes the text that is available: abstracts, full-text, or both.'
CREATE TABLE `publishers` ( `PublisherID` int(11) NOT NULL AUTO_INCREMENT, `PublisherName` varchar(200) DEFAULT NULL, `PublisherURL` varchar(245) DEFAULT NULL, PRIMARY KEY (`PublisherID`) ) ENGINE=InnoDB DEFAULT CHARSET=latin1 COMMENT='This table describes the Publishers that publish the journals. This will be a M:M relationship so will be connected to a bridge table called journals_has_publishers.'

CREATE TABLE `journals_has_publishers` ( `PublisherID` int(11) NOT NULL, `JournalID` int(11) NOT NULL, PRIMARY KEY (`PublisherID`,`JournalID`), KEY `fk_journals_has_publishers_journals1_idx` (`JournalID`), CONSTRAINT `fk_journals_has_publishers_journals1` FOREIGN KEY (`JournalID`) REFERENCES `journals` (`JournalID`) ON DELETE NO ACTION ON UPDATE NO ACTION, CONSTRAINT `fk_journals_has_publishers_publishers1` FOREIGN KEY (`PublisherID`) REFERENCES `publishers` (`PublisherID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=latin1

(Parent table: Organizations. Child table: Programs (non-mandatory))

CREATE TABLE `organizations` ( `OrganizationID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'pk for Organizations table', `OrganizationName` varchar(245) DEFAULT NULL COMMENT 'Name of the Organization', `OrganizationURL` varchar(245) DEFAULT NULL, `OrganizationInfo` varchar(1000) DEFAULT NULL COMMENT 'Describes how the organization is interacting with the Program.', PRIMARY KEY (`OrganizationID`) ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='lookup table for Programs. Only some programs are run by an official organizations.'

CREATE TABLE `programs` ( `ProgramID` int(11) NOT NULL AUTO_INCREMENT COMMENT 'PK of this table', `ProgramName` varchar(245) DEFAULT NULL COMMENT 'Name of the Program', `ProgramInfo` varchar(1000) DEFAULT NULL COMMENT 'Information describing how this program works in regards to accessing it''s journals.', `ProgramURL` varchar(245) DEFAULT NULL COMMENT 'the URL of the program, sometimes this can be the homepage of the collection or it can be a specific page within the collection site. Depends on how detailed the site is.', `OrganizationID` int(11) NOT NULL, PRIMARY KEY (`ProgramID`), KEY `fk_programs_organizations1_idx` (`OrganizationID`), CONSTRAINT `fk_programs_organizations1` FOREIGN KEY (`OrganizationID`) REFERENCES `organizations` (`OrganizationID`) ON DELETE NO ACTION ON UPDATE NO ACTION ) ENGINE=InnoDB DEFAULT CHARSET=utf8 COMMENT='Programs table describe the program that negotiates with countries for certain collections of journals. This table has a 1:M relationship with Organizations table (1 Org. can
participate in several Programs.) and has a M:M relationship with Countries table.