ICT 4014 CAPSON PROJECT REPORT

Effective Tools for Self-Archiving
in Institutional Repositories.

By:

Trey Chakulya (2017012983@student.unza.zm)

Elijah Chileshe (2017008345@student.unza.zm)

Sharon Kangwa (2017003527@student.unza.zm)

Royce Mwanza (2017013155@student.unza.zm)

Victor Mwelwa (2017012953@student.unza.zm)

Supervisor: Dr Phiri Lighton
Department of Library and Information Science
The University of Zambia
DECEMBER, 2021

Table of Contents

ABSTRACT

The introduction and implementation of institutional Repositories (IRs) has proven to be a success over the years and has helped scholars make their work known and available. The University Of Zambia (UNZA) like any other Higher education institution also has a repository which is used to preserve and make research outputs: pre-prints and post-prints and electronic thesis dissertation (ETDs) of the staff faculty at the university accessible. The repository at the University of Zambia has functioned well for a good number of years, but despite the increasing interest and encouragement in the use of the repository and recognition that the publisher policies of open access (OA) makes self-archiving achievable, staff faculty deposits in the repository still remains low primarily because most staff faculty are held back by the perception that self-archiving is extra work and this reluctance is downplaying the purpose of IRs. Nevertheless, every problem has a solution and the solution to this is to make self-archiving in IRs as easy as possible for staff faculty and this will be done by designing and implementing effective self-arching tools which will make self-archiving for the UNZA staff faculty a breeze.

CCS CONCEPTS

• Computer System Organisation •Software Testing- Debugging

KEYWORDS

Institutional repository, self archiving, SAT,

1 INTRODUCTION

Institutional Repositories are a specialized type of Digital Libraries implemented by Higher Education Institutions in order to store and make available scholarly research outputs. The University of Zambia (UNZA), like most Higher Education Institutions, has set up an IR ("UNZA Repository Home") which, in part, is used for archiving faculty staff authored pre-prints and post-prints, and Electronic Theses and Dissertations (ETDs). However, while UNZA has had a functional repository for a considerable number of years, the content that is in the repository or IR does not represent all the work that has been done by its staff and students in terms of writing. In addition, most of the recent work by the University of Zambia staff and student's academic papers do not reflect in the institutional (Zulu, 2019). This is enough to tell us that uploading literature or academic work on the IR is a major issue. Prior literature has identified a number of issues associated with the low uptake of IR content, some of which are lack of self-archiving practice, insufficient number Library staff available to deposit content and lack of IR awareness of faculty staff. One of the solutions that have been suggested involves the use of easy to use and effective self-archiving tools and it is the focus of this paper.

This project is meant to design effective third-party self-archiving tools that will improve the uptake of IR content, the main focus is on coming up with tools that will allow the specification of relevant metadata elements and are easy to use.

This is in an attempt to avoid the long process library staff avoid adding content on the repository also to make adding content to the repository fast , easy and effective, by creating tools which can be used on any platform for individual authors to self archive or to upload their content into the repository. This will be done by identifying the easiest and yet effective way of pushing data to the institutional repository so that the library staff is less burdened and the individual author does not go through a lot of trouble to upload their work.

2 BACKGROUND

Self-archiving is a strategy used by authors to make their scholarly works available on the open web and to provide open access (Phiri, 2018). Self-archiving was introduced as a means of making the publications of scholars readily available and accessible to a selected scholar community, not forgetting the storage of the same Scholarly work bringing about efficiency and effectiveness among scholar communities . Many institutions like the university of Zambia (UNZA) have Institutional Repositories (IR's) and do practice self-archiving however research undertaken by the team clear shows that the practice is not being fully utilized and implemented by both the Unza faculty staff and unza scholar community at large and the current state of the Unza repository clear supports this argument, the Unza repository has recorded low levels of digital content uptake and outdated scholar content.According to (Phiri, 2018) Most Universities face a problem of low uptake of Institutional Repository content due to the lack of self-archiving practice. The primary goal now as far as self-archiving and the use of IR at UNZA is concerned is to find ways of getting faculty staff and the UNZA scholar community at large to self-archive. This leads to the main objective of this research which is developing self-archiving software that will make depositing of digital content effective and easy which in turn increases efficiency. The team of Victor Mwelwa, Sharon Kangwa, Trey Chakulya, Royce Mwanza and Elijah Chileshe have developed self-archiving software tools that will make depositing of digital content easy and effective. These self-archiving software tools (SAT) make self-archiving easy, the tools have been designed to reduce the seemingly monotonous ways of self archiving, enable the users to easily deposit content into the UNZA repository, reduce the time it takes to deposit content making self-archiving easy, effective, efficient and fun.

2.1 Related Work

The libraries in the academic institutions play the role of the imparting knowledge to the academic fraternity of an institution since ages without discrimination for the dissemination of knowledge among student of different cultural heritage. However, during the past few decades, it became difficult for the academic libraries due to the scarcity of space. The emergence of IR is a new phenomenon that provides an opportunity for an institution to share its digitized intellectual wealth with the worldwide community of scholars and provide long-term preservation solutions.

According to [1], presently the IRs are shifting rapidly multi-dimensional in the academic environment however, its consensus is not yet clear among the repository managers, scholars, and the faculties. In this age IRs are seen as integral parts of every academic institution's infrastructure. During the last couple of decades, plenty of literature has been generated on the topic of IRs on different aspects. it has been concluded that despite the importance of institutional repositories there is some reluctance in uploading content in the repositories and some of the reasons are attributed to the less of man power in librarians, complexity of the system and not being aware of the existing of an institutional repository [2]. Hence the idea of creating third-party self archiving tools to solve these problems.

2.1.1 Self-archiving

Self-Archiving as a revolutionary way of publishing has been a myth for a long time. advocates have emphasized its advantages and significance in scholarly communication. Although having already found self-arching difficult to promote and manage, practitioners tend to pinpoint its potentials and take an

optimistic view. In most discussions, self-archiving has been considered to be an integrated part of the development and practice of digital repositories (Xia and Sun, 2007). However, Self-archiving is a broad term often applied to the electronic posting of content. It is also referred to as a strategy used by authors to make their scholarly works available on the open web. In this context, the contents are usually journal articles, conference or technical reports, theses and dissertations, or data sets. A scholarly work is self-archived if it is posted to a personal or professional website, deposited in an institutional repository, or contributed by the author to a disciplinary archive such as the Social Science Research Network (SSRN), arXiv, or PubMed Central. However, by self-archiving, researchers can make their scientific work freely and easily accessible to the world at large.(Wiley and son 2005) Self-archiving will result in research articles being accessible and significantly increase the article's visibility at UNZA. It will also save the researcher's time and effort, when an article is freely available and thereby, decreasing personal requests for the article. For example, Authors of articles published in(Wiley and son 2005), are permitted to self-archive the submitted (preprint) version of the article at any time, and may self-archive the accepted (peer-reviewed) version after an embargo period. Self-archiving is often referred to as Green Open Access. Furthermore, Self-archiving usually meets the openness criteria of research financiers. It is recommended for researchers to mention the openness of their previous publications in their funding applications. Therefore, Making the world a better place is always worth the effort! Every self-archived item brings one more research publication available to a great number of researchers who would otherwise be partially excluded from such scientific knowledge because of the high price of journals. At the same time, scientific knowledge becomes available to larger audiences, both nationally and internationally.Moreover, the introduction of self-archiving tools at UNZA will increase the uptake of content on the institutional repository.

2.1.2 Digital Libraries

Information technologies is an ever advancing field and the advances in Information Technology have revolutionized the role of libraries. Libraries are redesigning their services and how information is accessed by users in order to add value to their services and to satisfy the changing information needs of the user community. As a result, libraries face new challenges, competitors, demands and expectations. For example, one of the challenges the UNZA is currently facing is the low uptake of IR content into the Dspace repository. A digital library is a collection that contains different kinds of information that is stored in digital formats and accessible by computers. The content may be stored locally, or accessed remotely. The field of DLs is a multidisciplinary field that comprises disciplines such as data management, digital curation, document management, information management, information retrieval and li-brary sciences. According to [5], digital libraries provide and enable users to have access to a wide variety of information. Therefore, digital libraries provide users with different kinds of information that can easily be accessed anywhere at any time as long as the users have access to the internet. Digital Libraries are currently being used widely in various academic institutions around the world. Most of the information seekers whether researchers or not who use Google never go past the first page of results. The first page of results on the google website renders information that is trusted and mostly selected by various users. Library resources should match Google's ease of use and its criteria for first page listing. Library tools must exhibit all the qualities of what Ranganathan calls a 'well-arranged collection'.Digital Libraries are mostly used by researchers to gain access to particular information as well as upload digital content onto the Digital Libraries. Various institutions use different Digital Library Open source

software tools for storing their information. Different DL software's have different problems that come with them and different solutions have to be implemented in order to solve some problems users face with the Digital Libraries. Some of the different Digital Libraries Open Source Software Tools such as GreenStone Digital Library Software, DSpace Digital Library Software,Eprints Digital Library Software, FEDORA Digital Library Software and many others. However,The self-archiving tool which is to be implemented will be implemented in such a way that they will cut the tedious way of interacting with the institutional repository at the UNZA. For example, the only work the person will be doing is signing up or login then upload the content they wish to upload and the system does the rest. This will boost the uptake of content as many will be willing to upload content because the process will be easier to understand and do.

2.1.3 Proposed Solution

IRs being a specialized type of Digital Library enables information to be deposited and accessed by various users. Institutional Repositories are implemented by Educational Institutions in order to collect, preserve and disseminate digital copies of scholarly work created by faculty and students in individual Universities. Most Universities face a problem of low uptake of Institutional Repository content due to the lack of self-archiving practice [7]. This then leads to the Institutional Repositories having less information or lack of latest information that was published by faculty or students. Therefore, this project plans to accomplish the designing of effective third-party self-archiving tools that will improve the uptake of IR content. The self-archiving tools shall enable the users to easily deposit content into the repository, they shall also reduce the time it takes to deposit content and at the same time, the tools will be designed in such a way that they shall be easy to use by the users.

2.1.4 Problem statement

Currently, there has been low uptake of digital content on the IR. We need to have a clear way of publishing content on the IR in order to increase the uptake of digital content, so that the publication of content on IR will be efficient and easy, as authors will have the role of uploading content on their own. Content has been delayed on the IR, because we don't have enough visibility or the easiest way of publishing content on the IR. Therefore, we will implement third party self-archiving software tools which will enhance the visibility and ease of use to this process. Thereby, quantifying the uptake of digital content on the IR at the university of zambia (UNZA).

2.1.5 Aim

In the past years tremendous development in information production, acquisition and dissemination has been witnessed. This was possible through the deployment of various Open Access models of scholarly communication as part of concerted effort to weaken the effort of serial pricing crisis on the library and library users. one such model is IRs (self-archiving). However, the main aim of this project is to implement a software tool which will enhance the uptake of content in the IR at UNZA.

2.1.5.1 Objectives

2.1.5.1.1 General Objectives
1. To implement third-party self-archiving software tools that will increase the uptake of digital content in IR.

2.1.5.1.2 Specific Objectives
1. To quantify the uptake of Institutional Repository content in the University of Zambia Repository.

2.To design and implement user friendly, easy to use and useful repository software tools.

3.To evaluate the usability of the self archiving software tool.

2.1.5.1.3 Research Questions
1. How does the non-implementation of self-archiving strategies impact the uptake of content in IRs?

2. Is it possible to design and implement effective and usable self-archiving tools?

2.1.6 Significance of the Study
The creation of third party self archiving tools for the institution repository is inspired by the fact that a lot of faculty is reluctant to upload content on their own and this is because of the claim that it is not easy to navigate through the institution repository. hence the reason while this study has been undertaken the purpose of this study is to make uploading content easy and this tools is meant to achieve that by making uploading content in the institution repository easy, faster and effective by cutting on the processes or automating some of the processes that users go through while uploading content on the institutional repository. This tool will not replace the repository it will only help enhance effectiveness and it is not a follow up on any tools it is new at the university of zambia, furthermore the self archiving tools will be created in such a way that it will be flexible to allow addition of other features as need arises thereby increasing the number of staff uploading content.

2.1.7 Ethical Consideration
Ethical considerations in any research have been identified or specified as one of the most important parts of the research. Researchers have identified research ethics as the core aspect of research and the foundation of a research design, an integral part of the research that needs to remain at the forefront of the research work. The importance therefore of the research ethics cannot be understated. The University of Zambia (UNZA) like any other institution has an Institutional Review Board (IRB) that has properly structured set of research ethical guidelines that every researcher conducting research at the university must follow, the IRB is responsible for ensuring the safety of the human participants and prevents the violation of their human rights. The IRB at the university reviews the methodologies and aims of the research study and ensures that the ethical guidelines are followed and if the conditions of the IRB are not fully met the research study will have to be amended. The following are the ethical guidelines followed in this research:

1.6.1 Informed consent

Informed consent is a major aspect of research ethics that demands that all participants must be adequately informed before taking part in the research with that said, the staff faculty during this study will be given sufficient information about the research and the purpose of the research in order that they make informed decisions. The staff faculty will only take part in the research when their consent to participate in the study is given as research ethics requires that an agreement between the participants and the researcher is made before the participants can participate in a research study and participants will also be informed that they are free to withdraw from the research when they feel a need to.

1.6.2 Voluntary Participation

Participation in all research is voluntary, therefore the sample target which is the staff faculty at the University of Zambia (UNZA) will by no means be forced, persuaded or deceived into taking part in the study. All staff faculty at the UNZA will be invited to participate in the study with a clear understanding that they are under no obligation to do so and that there will be no negative consequences for them if they do not assist in the research. If among the staff faculty there is a participant who cannot properly read or write and he/she does not understand what they are being asked to do and does not comprehend that the study being conducted is voluntary. In case of such a scenario research ethics for protection of the respondents or participants even from unintentional harm.

1.6.3 Respect of Confidentiality and Anonymity

The information supplied by the staff faculty and their answers will remain confidential and anonymous and will by no means be shared to the public or shared within the institution. Data collection techniques of ensuring the confidentiality and anonymity of the staff faculty will be implemented so as to hide their identity and response. For example the use of surveys and questionnaires.

1.6.4 Risk of Harm

Potential of harm to participants can be sometimes difficult to predict especially in a research study as this one where the risks are close to zero, however all researchers whether (student, professional, or academic) are well intention but the possibility that interaction with the participants may inadvertently harm them in some unintended way cannot be overlooked and therefore measures to counter such incidents have to be put in place and it is therefore the responsibility of the researchers conducting this study to make the risk to benefit ratio minimal and if the possible risks during this study over weigh the benefits the research design will be abandoned or redesigned. The possible risk of harm identified during the research are: psychological and social harm. However due to the nature of this research study the potential of harm to the faculty staff at UNZA is minimal.

1.6.5 Communicating of Results

The communication of results after research is done will be judged on three: plagiarism, academic fraud and misrepresentation of results.

3 LITERATURE REVIEW
3.1 Introduction

The awareness and significance of self archiving among the scholar community is rapidly growing in spite of the fact that the self-Archiving practice has been around for many years and has greatly contributed to the publication of Scholarly work and has also helped many scholars receive the recognition they deserve. (Cerejo, 2013). The benefits of practicing self-archiving include increasing visibility, readership and citation of one's publications and increasing researchers' recognition in the field. However the gaps and holes in the practice prove that there is still plenty of room for improvement and this improvement can be achieved by the development of tools that make the practice easy. In an effort to examine the trends of creating self-archiving tools that are effective and easy to use, some studies related to the topic were reviewed and the outcome are here below.

According to Harnad (2001) What was needed was something that would make the fruits of distributed, institution based self-archiving equivalent to those of centralized self-archiving, and the key to that was to introduce and agree upon metadata-tagging standards that would make the contents of all the distributed archives interoperable, hence harvestable into one global "virtual" archive, all the papers searchable and retrievable by everyone for free, without having to know in advance where they happened to be individually archived, or in what form. The Open Archives Initiative (OAI) has provided the meta-data tagging standards and a registry for all OAI-compliant Eprint Archives, and the Self-Archiving Initiative has provided the free software for creating OAI-compliant Eprint Archives, interoperable with all other Open Archives, ready to be registered and for their contents to be harvested into searchable global archives. Distributed Institution-based self-archiving is the natural way to generalize the practice of self-archiving across disciplines and institutions. It is not only the author who benefits from research impact. He observed that the reason promotion and tenure are contingent on research impact is that funding is contingent on it too.

Ahmad and Mohammad (2017) focused on improving the functionality and effectiveness of an IR, outlined the importance of assessing its activities at specific intervals. He suggested that IR managers can check the usage statistics like actual page view, downloading of content, traffic sources etc. Page view by countries and search engines will help theIR managers to evaluate its visibility and acceptance. IR managers can conduct a user study to know the opinion of stakeholders about IR and collect their recommendation in order to uplift the services of IR. Input activity and usage are inextricably linked, that is scholars are more likely to use an archive if it has significant input activity, and on the other, they are more likely to deposit their work if an archive is highly used, thus providing greater visibility to their research (Shearer, 2003). The ability to track citations and analyze the contributions of repository content to scholarship is an important incentive for scholars. The combination of input activity, usage, and citation analysis will give the full picture of repository effectiveness

Swan and brown (2005), in their effort to find out how many people were self-archiving carried out a research, the question permitted respondents to check any number of options, so that some of the people who have self-archived on their websites, for example, may also have self-archived in institutional or subject-based repositories. The overall findings presented did not reveal the true level of self-archiving within the population. To find out what this was, the results for this question were analyzed on a case-by-case basis so that the number of individuals represented in the results was revealed. The total number of people in the respondent population who have self-archived in any way is 631, that is, 49% of the total population. This total population of 1296 researchers has carried out 1303 individual acts of

self-archiving; since 631 individuals have self-archived, the average number of self-archiving acts per self-archiver is 2.1. The last time they measured the level of self-archiving 1, 2, 10 months earlier than this present survey, 23% of the population had self-archived in some way at least once. The proportions of self-archivers in each of the sub populations surveyed is: 'Interested and informed' population: 35%, Archived population: 85%, Southampton population: 77%, Randomly-selected population: 42%. It can be noted that the 'interested and informed' population is not self-archiving as much as the average, or as much as the randomly-selected population, so the term 'interested' is something of a misnomer. The results for the 'archived' and 'Southampton' populations are also worthy of comment. The archived population comprised individuals whose names were obtained from papers stored in open access archives and in practice, every one of these individuals should be a self-archiver, by definition. Nonetheless, 15% of them have not recorded any self-archiving activity in this survey, underlining the fact that an author may become a self-archiver by default, that is, by his or her papers being archived by co-authors or by proxy archivists (librarians, archive administrators, colleagues). In his conclusion he emphasizes the importance of self-archiving and the need to encourage faculty to practice self-archiving.

Grundmann (2008), explored the reluctance of faculty to self-archive due to the idea that it takes too much time and it is too complicated to do. He described the "learning curve" associated with the institutional repository as a deterrent. He further notes that there are other issues involved in the amount of time and difficulty it takes to self-archive publications, including investigating publisher permissions, digitizing paper documents, and ensuring privacy for certain kinds of research data. In addition, sometimes a few extra minutes are a few too many. A third concern is the quality of material housed in a repository. There is a perception that repository contents are not peer-reviewed and that repositories conflict with the prestige factor of publishing in respected journals for promotion and tenure decisions. This, according to Grundmann, suggests confusion over the purpose of IRs, including the misconception that they are a replacement for traditional journal publishing.

Concluding in his paper, he notes that the solutions to low faculty participation in self-archiving in institutional repositories are complicated hence adds that the most common finding with each strategy for increased participation is that no one approach will work on its own. The most promising approaches include promotion of self-archiving in repositories through targeted education and outreach, combined with faculty-supported mandates, and increased findability through a greater web presence. Librarians and other IR supporters must take care that their approaches are appropriate for their institutional culture. Studying local faculty needs and preferences for communication channels will help in creating a more useful outreach strategy.

Mandates are most useful when the faculty at the institution are supportive. And finally, increased awareness of the repository can come through an increased web presence on the portions of the institution's website that are most likely to be seen by faculty. The most important consideration is that the IR meets the needs of the faculty as well as their institution. A journal checker tool to enable their authors to provide the right information before uploading journals to the institutional repository. According to (), the interface was designed such that it will provide an author with concise information regarding the correct requirements of the journal or everything that a journal article should have before it is uploaded to the repository. Over time, the tool will need to provide guidance as to whether the

journal meets all the requirements: the basic mandatory conditions for all publication venues and all the mandatory technical conditions for all publication venues, as specified in the Technical Guidance and Requirements.

If the journal is not aligned to requirements the tool should suggest alternatives.Upon the return of results, the user should see very clearly if the journal they searched for meets the "initial requirements'' or Upon the return of result an alternative should be given. This tool focused on journal articles only and did not focus on all scholarly works hence the gap to make sure that all papers regardless of the the type are detected and checked before they are uploaded so that it is easy and faster to upload content.

The purpose of this review was to view the trends in self-archiving studies within the past years and see how self archiving improves the deposit of content on institutional repositories. It is clear from the reviewed research that there is a need for faculty member's to practice self-archiving in Universities, so as to improve the deposit of content on the IR's. Along with this it is also self-archiving which would reduce the time and processes taken for content to be deposited on IR. More research and testing is required for self-archiving tools, to gain better understanding of what would be the effective way of increasing the uptake of digital content on IR. This field of inquiry is very important as at its center of concern with helping publishers of content find effective means of depositing content on IR.

4 METHODOLOGY
This section describes the chosen research methods, and it evaluates the studies overall validity and reliability. It specifies procedures or techniques used to identify, select, process, and analyse information about a topic. This section states how the data was collected and how it is going to be analyzed in response to the research questions, the research design, study site, sample size, measurement instruments, data analysis tools, sampling technique, as well as the limitations encountered during the study. It will also indicate the methods used in coming up with the software and technology used.

4.1 Research Design
Agile approach was used for this project management, Agile is an iterative approach to project management and software development that helps teams deliver value to their users faster and with fewer headaches. Instead of betting everything on a "big bang" our team delivered work in small, but consumable, increments. To correct data from the stakeholders an experiment was set up. Questionnaires as well as interviews were used to collect data.

4.2 software implementation methods
There are many techniques prior to coming up with a software such as the agile, waterfall model as well as the rapid application development. However, for this project, we decided to use the combinations of these techniques. The combination of techniques was chosen due to their efficiency in balancing when resources are not enough.

4.3 Factors Influencing the uptake of digital content onto the University of Zambia Institutional Repository

Factors that influence the uptake of digital content on UNZA IR were identified by using the data collected from an online interview session. The interview was conducted using Google Meet. The reason why the online interview was preferred to the physical one was because of the COVID-19 outbreak which led to closure of the University hence in order to bring together all team members and other identified stakeholders, who happened to be in different places, online interview was appropriate.

4.4 Sampling Method

Sampling is the process of picking a sample out of the population to help with managing the resources as well as for the accurate results of the study. In this study a Non-Probability Sampling Method was used. There are a number of non probability sampling methods but for this study a convenience sampling was employed. Convenience sampling is easier to apply , because participants are selected based on availability and willingness to take part.

4.5 Measurement Instruments

Since this is qualitative research, Interviews were used to collect data through the use of Google Meet concerning the tedious work involved in depositing content on the UNZA IR. . Interviews were used to collect primary data from content publishers in UNZA IR. Closed and open ended questions were used because they enabled us to collect more detailed information. This tool was very appropriate because it enabled us to record the interview, which gave us room to play it back at any time.

4.6 Data Collection

Stakeholders were identified to be included in this phase of research. These stakeholders are: experts, non-experts as well as the library and information science department. These stakeholders were classified into two main panels: an expert and non-experts panel. The non-experts panel includes all the stakeholders except the experts. This classification was made according to their different nature and their different experiences. Therefore, data was collected from two lecturers and fourth year students from the lis department. This was archived by sending out online questionnaires as well as the use of google meets to conduct the interviews. These instruments were chosen because they are easy to use and suitable for this as we are currently living in a covid 19 era.

4.7 Data Analysis

Qualitative data analysis software tools were used to compare the effectiveness of the tools developed. There are a number of tools that could be used to analyze data, but for this research study AttrakDiff and Audio narrative were used. These tools helped targeted people to be more involved and closer to the data. The two proposed tools provided the right questions or tasks to help get the much needed information. This information will help in the measure of how effective the tools are.

4.8 Research Approach

The research method can either be qualitative or quantitative in nature, qualitative and quantitative research has been discussed by many authors for example Maxwell, 1998; and Thomas, 2003, pointed to the key points for a qualitative research design. In simple basic terms a qualitative research approach can

be defined as the process of collecting and analyzing non-numerical data e.g., text, video, or audio to understand concepts, opinions, or experiences. It can be used to gather in-depth insights into a problem or generate new ideas for research. With the above definition of qualitative research, our team decided to follow this approach as it is suitable for our project. In this research study, the key elements are to gather information through interacting with people who have experience in this field of our study. The study is aimed at developing effective tools for self-archiving in Institutional Repositories.This aim will be achieved through the objectives highlighted in the beginning of this paper.

5. DESIGN
5.1 REQUIREMENTS ANALYSIS
5.1.1 User Interview
An interview was conducted to gather feedback from the University of Zambia Library Staff about our proposed system. The Library Staff are the people that are responsible for the uploading of digital content to the Institutional Repository. The interview included questions that enabled us to gain an understanding on whether our proposed system would be helpful to the UNZA Library Staff and also to help us refine our initial requirements.

The interview was scheduled with one participant who is the one in charge of the synchronization process with the IR. After the interview was conducted, we concluded that the system we had proposed would be useful in that it would make the depositing of records to the UNZA IR easy and less time consuming compared to the traditional way.

5.1.2 Use Cases
Use case diagrams for the Administrators, Lecturers and the Postgraduate Students were drawn to help define the goals of the users of the proposed application to be developed, which is the SAT application. Once the goals of the users were identified, we began to implement the functionality.

The goals of the Administrators and the Lecturers as shown in Figure 1 below are separated by the tasks they perform. The Lecturers are able to login to the SAT application, they are able to upload content to the system which includes the process of uploading a bitstream and also assigning metadata to a particular record being uploaded. The lecturers are also able to review all the records that have been uploaded by the Postgraduate Students as they wait to be synchronized with the Institutional Repository as well as give feedback to the students. The lecturers can also delete content from the SAT application that they do not want to be uploaded to the Repository. On the other hand, the Administrators are able to carry out all the tasks that the lecturers can perform but what differentiates the Administrators from the Lecturers is that the Administrators are able to carry out the process of synchronizing content with UNZA Institutional Repository as shown in Figure 1.
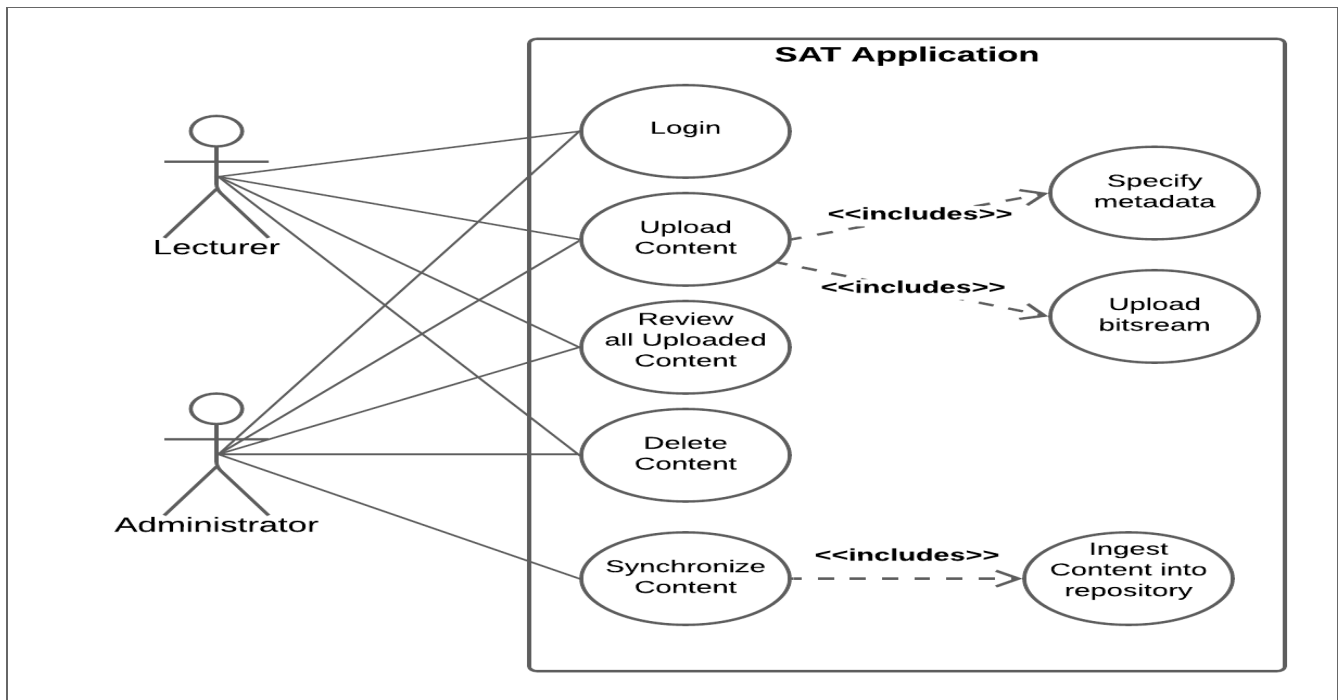
Figure 1. Use Case Diagram for the Administrators and Lecturers

Other users of the SAT application are the Postgraduate Students who are able to upload records to the UNZA Institutional Repository prior to them graduating. They are able to login to the application, deposit content, review the content that they have uploaded as well as delete the content each student has uploaded individually as shown in Figure 2 below.
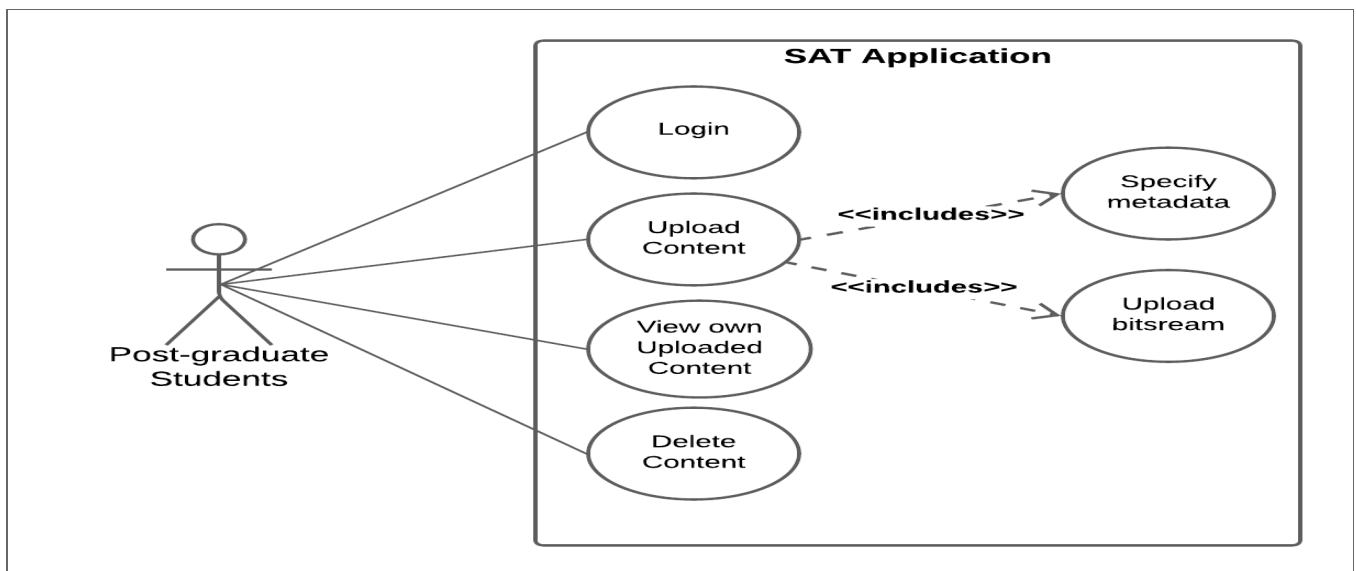

Figure 2. Use Case Diagram for the Postgraduate Students

The process one goes through when interacting with the SAT application for the Administrators, Lecturers and Postgraduate Students is shown in Figure 3 and Figure 4 respectively. The administrators and lecturers log in to the application and review all the uploaded records by the Postgraduate Students. After reviewing the records, the administrator can then go ahead and synchronize the records with the UNZA Institutional Repository as shown in Figure 3 below.
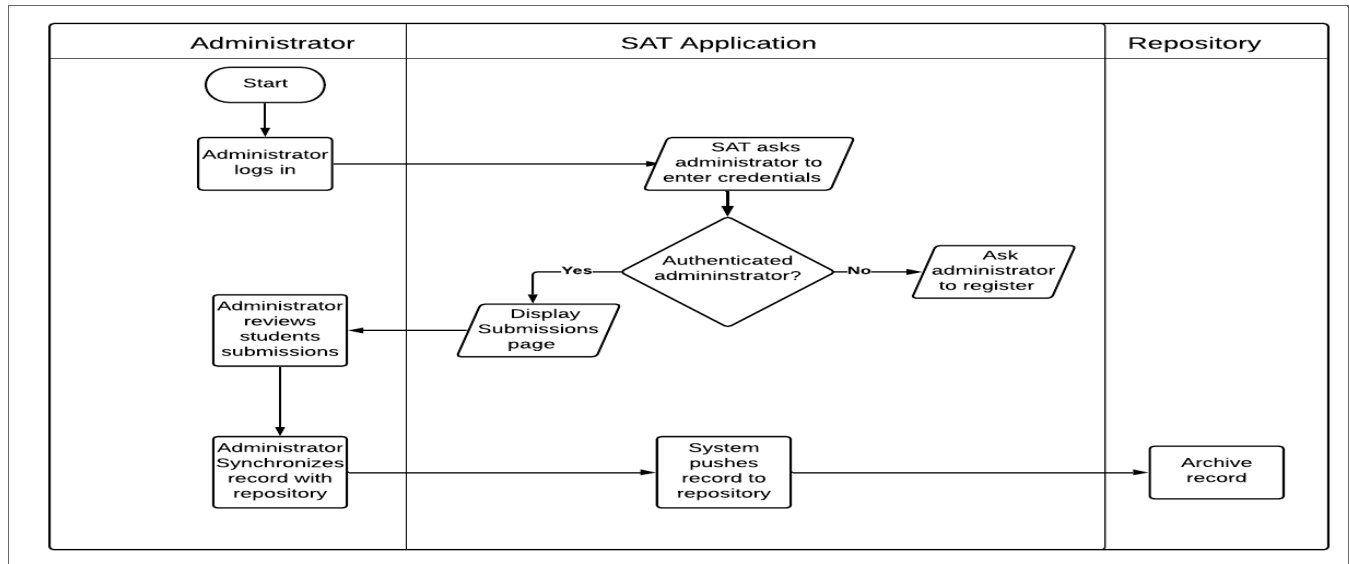
Figure 3: Administrator and Lecturer Process Flow

The process that the Postgraduate Students go through when uploading records to the SAT application such as logging in, selecting the PDF to be uploaded and specifying the metadata of each record is shown Figure 4 below.
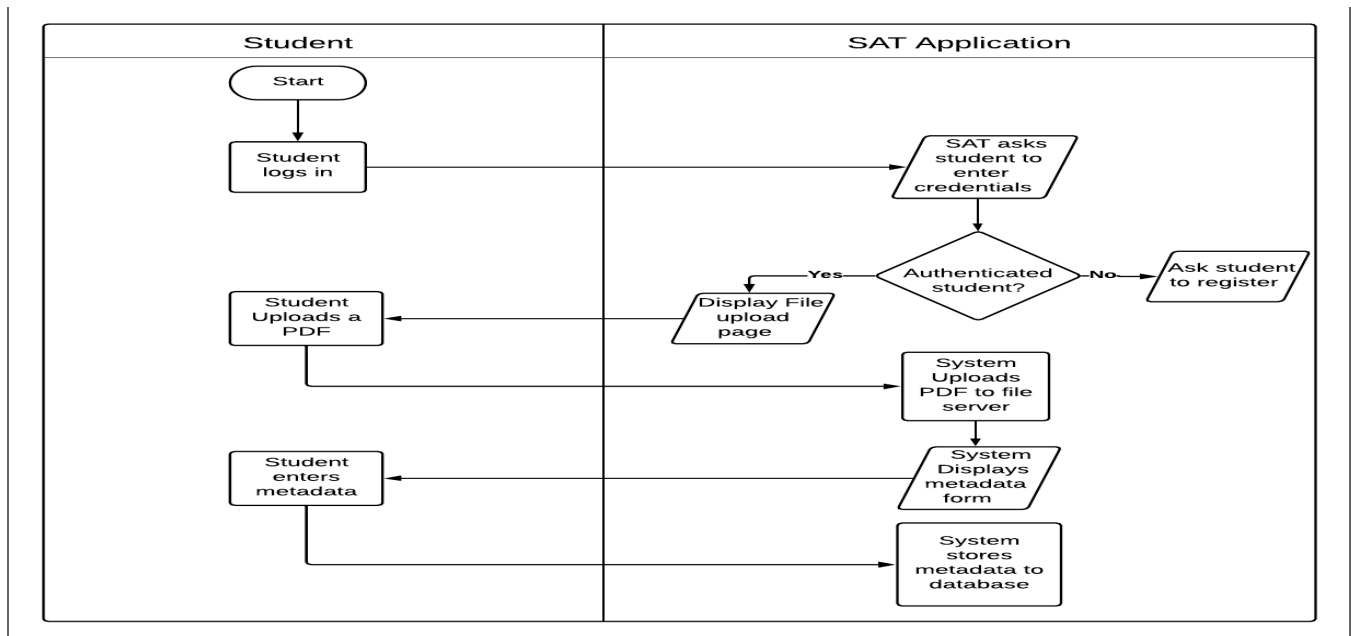
Figure 4: Postgraduate Student Process Flow
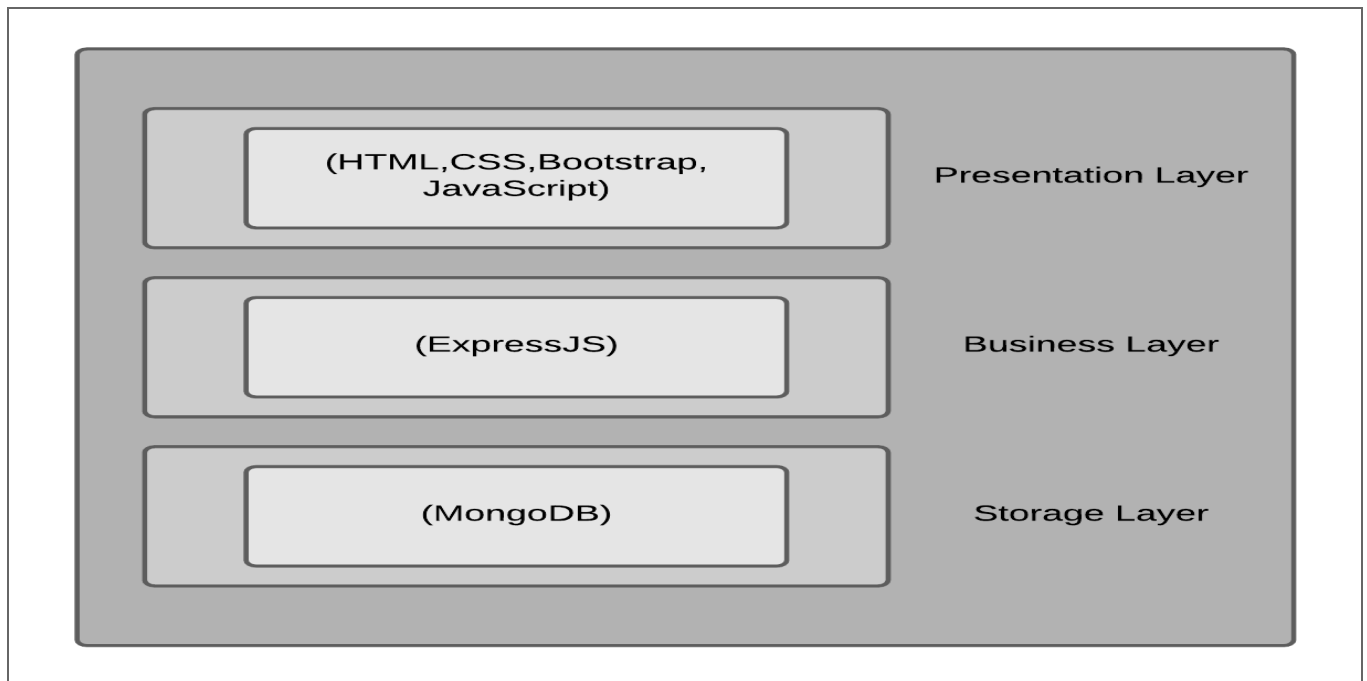
Figure 5: Layered Architecture of the SAT application

6. IMPLEMENTATION

This chapter presents the actual implementation of the Self Archiving Tool (SAT) and shows various functionalities associated with it, which can be useful in the increase of digital content on the UNZA IR. it is also an effective articulation and description of the design and implementation of the self archiving software tools.

The Self-Archiving software tools will be web based applications designed using the MERN STACK technology, the intended purpose for the tools is to help improve the uptake of digital content to the Institutional repository (IR) at the university of Zambia (UNZA) by reducing the seemingly long processes of uploading content to the IR and changing the monotonous way of interacting with the IR thereby make uploading content easy, effective and user friendly. The diagram below gives a graphical description of the design of the tools and processes:
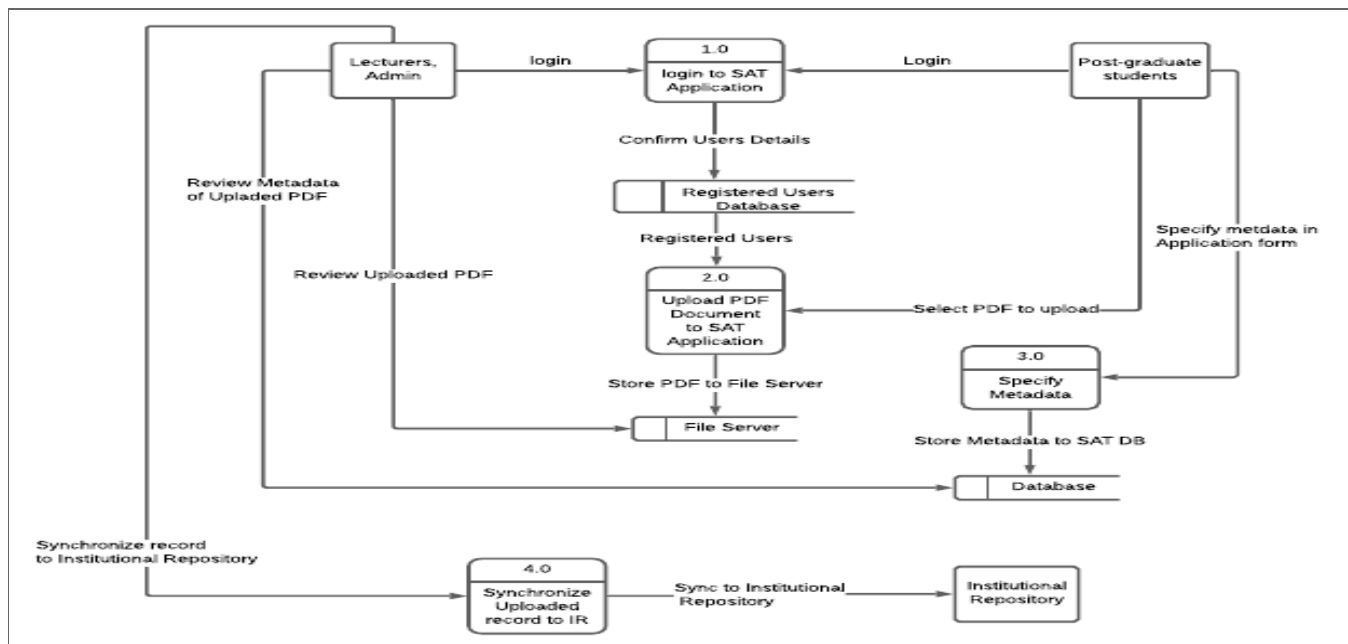
Figure 6: Data Flow Level 0 diagram showing the ingestion process into the SAT Application and to the IR

The Self-Archiving software tool to be implemented is going to be a web based application through the use of MERN STACK technology, the tools are meant to change the monotonous way of interacting with the IR at UNZA which will inturn reduce or cut down on some of the long processes involved. For a web based application to be successful or called a success depends on the applications ability to provide the suitable facilities required to deal with the task at hand, a task that should be executed in such a way that it enables users to fully and effectively exploit the system. The only way a user can exploit a system and use it to his or her advantage is by interacting with the system and the only known possible way of initiating interaction with a computer system is through the various interfaces of the system.
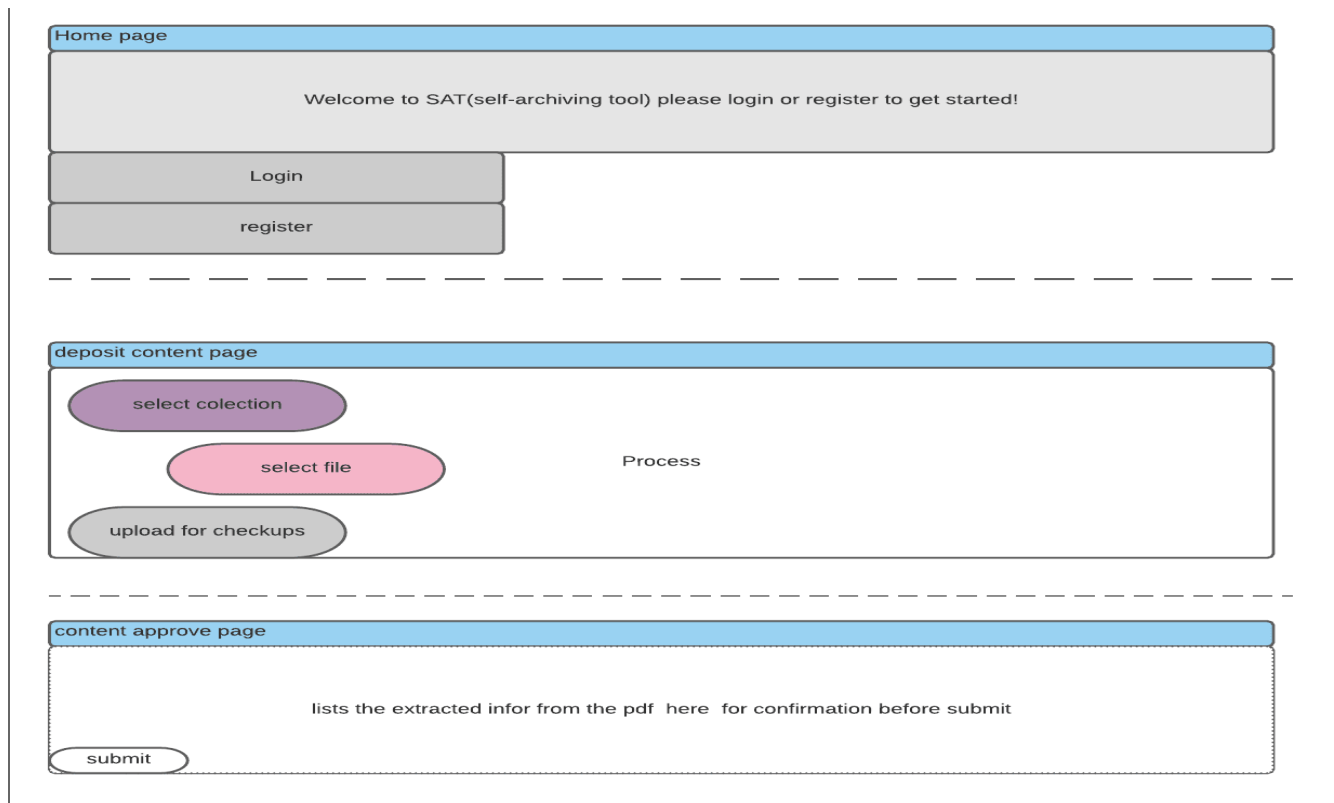
Figure 7: Diagram showing the SAT application as a whole

The user interface will have a sign in and login buttons, the buttons will be used for either signing up if not registered or logging in if registered. After the user logs in he/she will will see the page or button which requests to deposit content, an unregistered user that initiates interaction with the tools for the first time should be able to see a sign in page after opening the application and after registration the user should be able to log in, explore the application and use the application for its intended purpose. The diagram below explains further on the user interface of the tools.

The user interface does not just enable users to interact with the system but it also has a persuasion side to it, it lures users in, into wanting to know more about the system it makes them feel in control which makes them to want to explore and learn more about the system and the operations of the system and this is mainly where graphical features come in. The graphical features shown in the diagram above are what primarily attract usage from the user. These features enable the end user to follow and execute instructions and also helps the user get familiar with the functions of the system.

6.1 Development Frameworks

6.1.1 SAT application

The application was developed using the MERN Stack and it was designed to work with DSpace version 6.3.

6.1.2 Backend

The backend of the SAT application was developed using ExpressJS which is a backend development framework for NodeJS. ExpressJS is used for developing the backend of web applications as well as the Application Programming Interfaces (APIs).

6.1.3 Frontend

The frontend or user interface of the SAT application was developed using JavaScript, HTML 5, CSS 3 and Bootstrap. JavaScript and HTML 5 were used to display the text on the Application while CSS 3 and Bootstrap were used to style the different HTML elements that were specified.

6.1.4 Database

The database that was used for the development of the SAT application was MongoDB. It is a document-oriented database and is referred to as a NoSQL database. During the development process, MongoDB Atlas was used which is an online cloud database. Another type of database that was used during the development process was MongoDB Compass which was also used for querying the database and while MongoDB Atlas was used online, MongoDB Compass on the other hand was required to be downloaded to the local machine in order to be used.

6.2 Development Environment

6.2.1 Integrated Development Environment

The Visual Studio Code integrated development environment (IDE) was used when developing the SAT application.

6.2.2 Version Source Control

During the development of the application, the source code was required to be managed throughout the project. Therefore, Github which is a code hosting platform for version control and collaboration was used to carry out the management of the source code.

6.3 Hardware Interfaces

Since the tools being designed will be web based applications it means they will have to run on devices that meet certain functional requirements and most have access to the internet, devices such as: mobile

smart phones and a windows or Mac device, the devices should also support languages such as HTML and JavaScript.

6.4 Communication Interface

The data that will be transferred between the server, the schools database and the specified users' devices shall use the networking protocols: Transfer Command Protocols (TCP) /Internet Protocols(IP) that will be executed through a secured connection. This is to ensure prohibition of unauthorized access.

6.5 Trello Board

A simplified trello board was created in order to track the software development process of this project. The trello board consisted of five (5) lists which were; "To Do", "To Do (Internal Deadlines)", "Pending Tasks", "Completed Tasks" and "Uncompleted Tasks". As new tasks of the development of the software were identified, they were written onto cards that contained details of the individual tasks to be performed as well as the starting date and end date during which the tasks were required to be completed. The tasks were first stuck on the "To Do" list. During the task's development, they were then moved to the "Pending Tasks" and once the tasks were completed, they were moved to the ""Completed Tasks" list. On the other hand, if the tasks were not completed, they were moved to the "Uncompleted Tasks" list.

6.3 Frontend Graphical User Interface
6.3.1 Registration Page

This is known as a registration Page. This is where the new user has to register before accessing the system. A critical look at this form, it has no menu as shown in figure 8 below.

Figure 8: Registration page

6.3.2 Login Page

This serves as the first interface that welcomes the user to the site, this page has no menus but from the page, user will be able to access the login form to login into the system as shown in figure 9 below, if the user is a first time user then he has to register as he is required to login with the username and password.
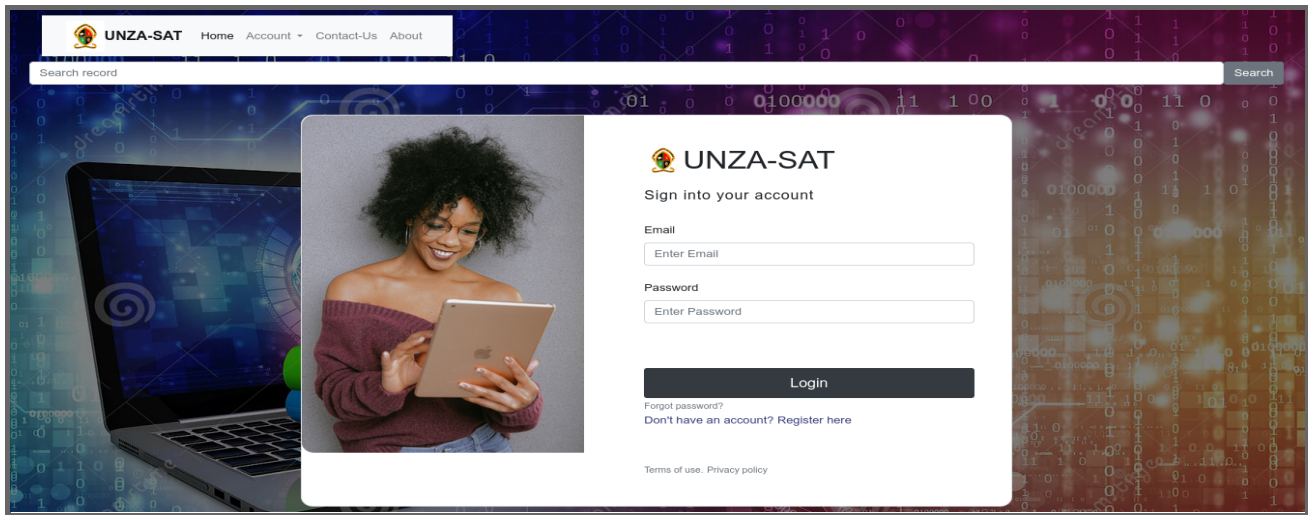

Figure 9: Login page

6.3.3 File Upload Page

Once the user successfully logs in to the application, a form is presented to the user that offers an opportunity to deposit content shown in figure 10 below.
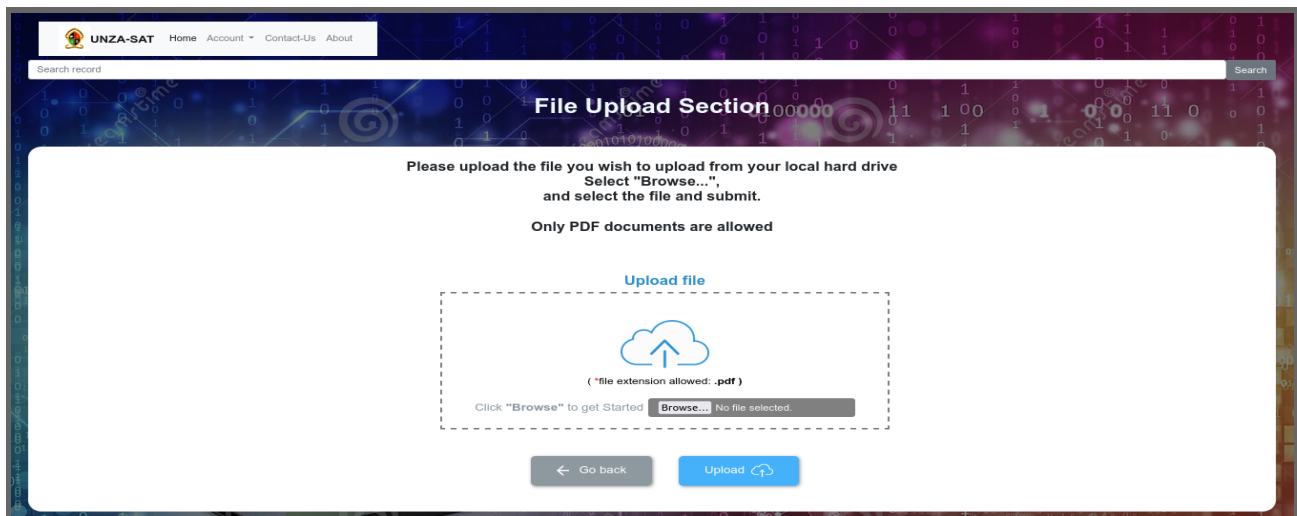
Figure 10: File upload page

6.3.4 Metadata From Page

This page is rendered when the user successfully uploads a document from the file upload page. The form page consists of a form with few questions that the user needs to enter in order to submit the metadata associated with the file that was initially uploaded. The form page shown in figure 11 below:

Figure 11: Metadata form page

6.3.5 View Submissions Page

The submissions page is displayed to the user once a record has been successfully deposited to the SAT application. The submissions page consists of a list of records that have been deposited on the application. The submissions page shown in figure 12 below:

Figure 12: View submissions page

6.3.6 View Individual Users Page

The view individual users page is a page that displays all the metadata elements entered when the users submitted their records. The submissions page is shown in figure 13 below:
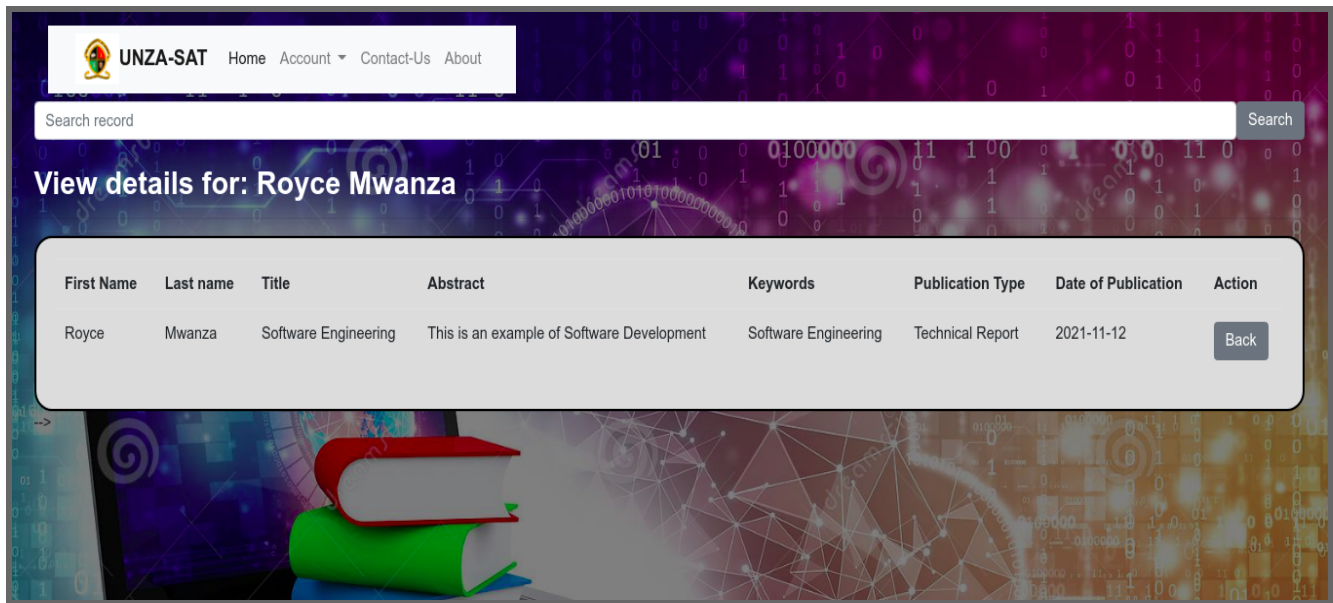
Figure 13: View individual submissions page

6.3.7 Edit Submissions Page

The edit page is a page that consists of a form that enables the users to edit some metadata fields that were entered initially when submitting the form to the SAT application. The submissions page is shown in figure 14 below:

Figure 14: Edit submissions page

7 RESULTS AND DISCUSSION

7.1 To quantify the uptake of Institutional Repository content in the University of Zambia Repository.

The UNZA IR holds about 6700 documents as digital content, this content is deposited by various schools which are divided into departments. However , in order to quantify the uptake of IR content in the UNZA IR, content was extracted from the IR using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) Protocol. The content that was extracted was in an Extensible Markup Language (XML) Format and it had to be processed in order to understand it. The content in XML format was processed using Python and the output was converted to Comma Separated Values (CSV) for Excel. The output in CSV files was then analysed using Google Sheets to come up with Data Visualizations in the form of Bar Graphs and Pie Charts.

### 7.1.1 Results for depositing of content in the IR per School

Figure 15 indicates the number of schools at UNZA and the number of content at least each school deposited from 2000 to 2013, findings shows that among the 13 schools which exist at UNZA the school of Education has the highest number of deposited content on the IR, followed by the School of Agriculture and Library then the school of Humanities, University Collection, Natural Science, Veterinary Medicine, Institute of Economic, Engineering, Mines, Institute of Distance, and the school of Law shows to have the lowest number of deposited content on the IR. However, the gaps which exist between the highest and lowest and the highest and the second schools to deposit content are big . This shows that there has been less contribution in terms of depositing content by various schools at UNZA.
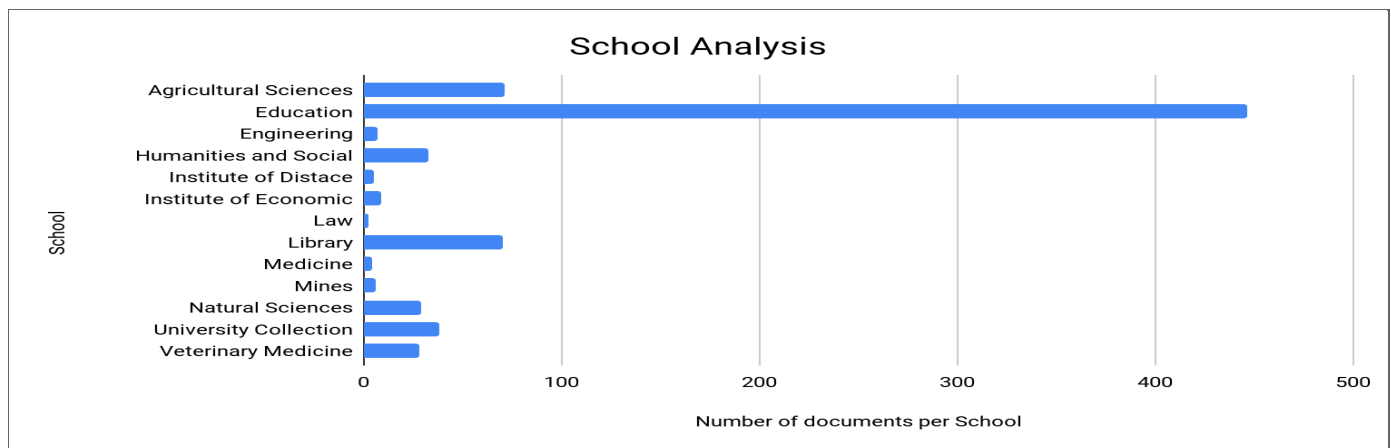
Figure 15 : Schools

7.1.2 Results for depositing of content in the IR per Department in the School of Education

Figure 16 indicates the number of Departments under the school of Education and the number of content at least each department under the school of education deposited from the year 2000 to the year 2013, findings shows that amongst the 8 departments which exist in the school of education, Language and social education department has the highest number of content deposited on the IR, followed by education, Library and information, educational, Adult education, Religious Studies, Mathematics and then the Primary education department has the lowest number of content deposited on the IR.
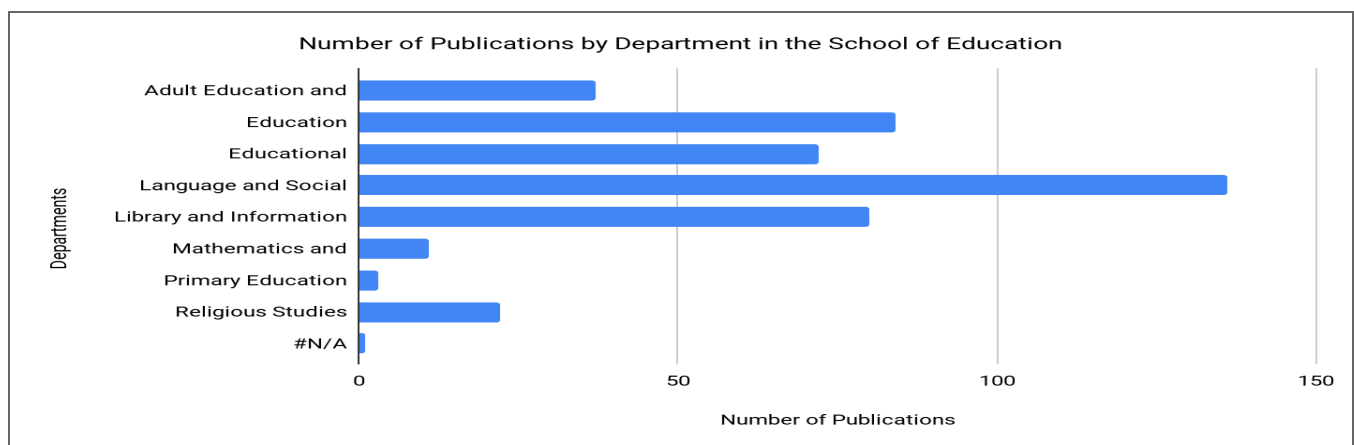

Figure 16: Departments

7.1.3 Results for type of content deposited to IR

Figure 17 shows the type of content that is deposited in the repository ory and how much of that content is uploaded according the the figure above it can be seen that articles have have the highest ingestion

rate with the 75.7% of everything that that is in the repository followed by book chapters which have 9.1% very far from each other.   The other section of the content has 3% the rest share a very small percentage of all the content uploaded in the repository.
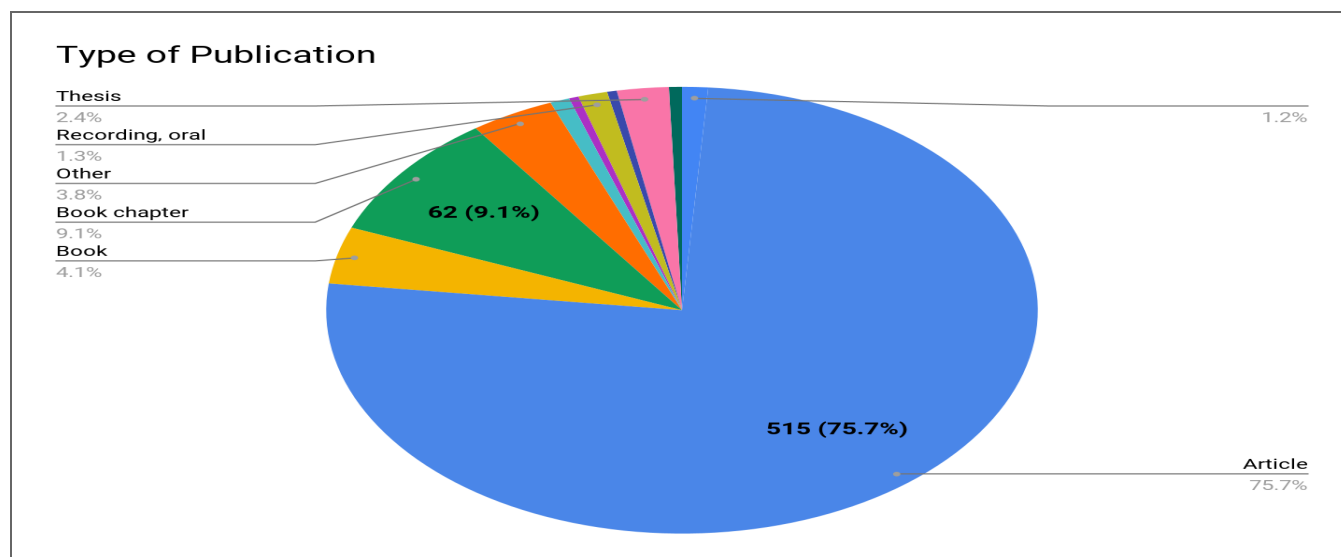
Figure 17: Type of Content

7.1.4 Results for trends in upload of content to the IR

Figure 18 shows the trends in the upload of content from the year 2000 to 2013. From 2000 to 2006 it can be seen that there was very low uptake of content in the institution repository. Less than 20 percent of content was uploaded in the repository. 2007 shows some changes because the numbers began to increase as a lot of faculty became aware of the repository the numbers continued to increase with 2011 being the highest in that period and followed by 2012 and then 2013. This also shows that there is no consistency that can help tell if the numbers will increase in a certain year or they will reduce hence it is difficult to predict. The university of zambia being the highest learning institution in the country ought to have an institutional repository that gets better with time because new schools are formed and new department are created and as years go by some schools grouw and hence enroll more undergraduate students and post graduate students hence the digital content or the academic content also is expected to increase.
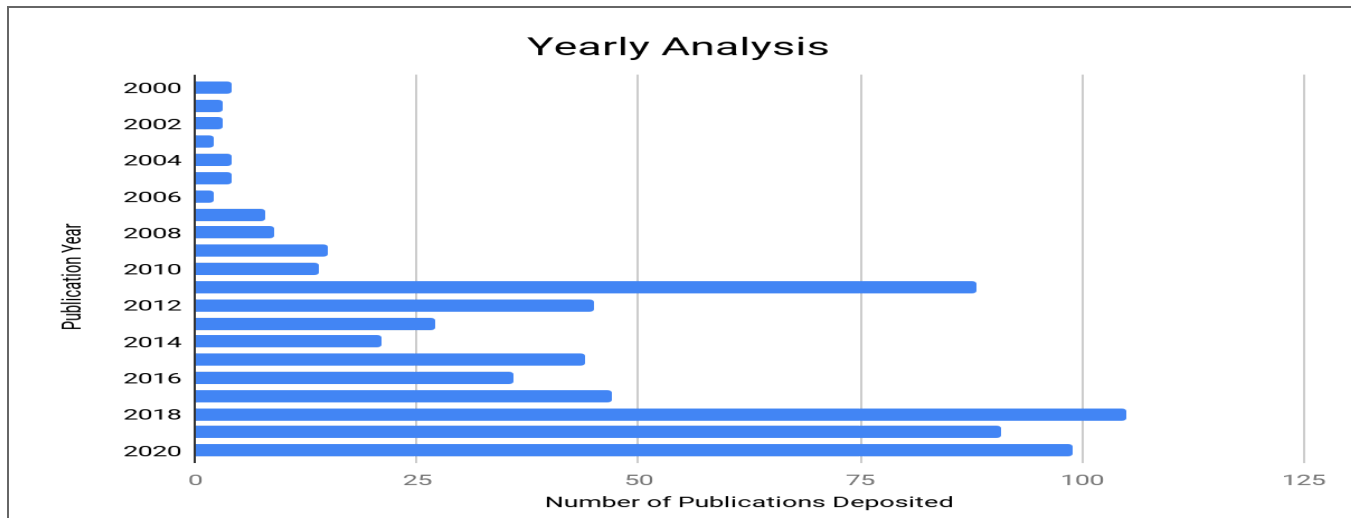
Figure 18: Year of Publication

7.2 To Evaluate the Usability of the Self Archiving Software Tool.
### 7.2.1 Usability Evaluation

When we completed the basic functionalities of the SAT application, we performed usability evaluations on it. To evaluate the application, we decided to use the Technology Acceptance Model (TAM) evaluation method. TAM suggests that when users are presented with a new technology, a number of factors determine the decision about how and when they will use it. The model includes two main factors influencing individual use which are perceived usefulness and perceived ease-of-use. The TAM is relatively easy to use and can also be used on small size participants. TAM provides a list of 12 questions that test the user's experience with the application. Each question requires that the user selects their answer on a scale ranging from strongly agree to strongly disagree.

7.2.3 Pilot Study

We conducted a pilot study of the SAT application in readiness for evaluation with the students. In the pilot study, we had our team members and colleagues use an incognito window to access the guidelines created for the applications. They were required to follow the guidelines that included a link to the PDF to be downloaded, a link to the application as well as metadata needed to input to the application during the depositing process of a record. From the pilot study, we were able to refine our guidelines as well as fix some errors that were identified on the application.

7.2.4 Evaluation with students

We sent out emails to the Bachelor of Information and Communication Technologies mailing lists for students to access the email and participate in our study. We received feedback from 5 students. The evaluation of the application with students took place remotely. We sent a description in the email explaining to them what our project is about as well as links to the applications. We had a list of tasks for the students to perform on the application. The tasks involved downloading a PDF document sent in the guidelines, accessing the application using the provided link, uploading the downloaded PDF document as well as specifying the metadata and also indicating the starting time and ending time of the submission process on the application. After they completed the tasks, they were required to fill in the TAM questionnaire that included Demographic Details and the starting and ending time it took for them to complete the uploading of the record to the application. **Figure** shows the questionnaire used during the evaluation. The SAT application interface users interacted with is the one in **Figure.**
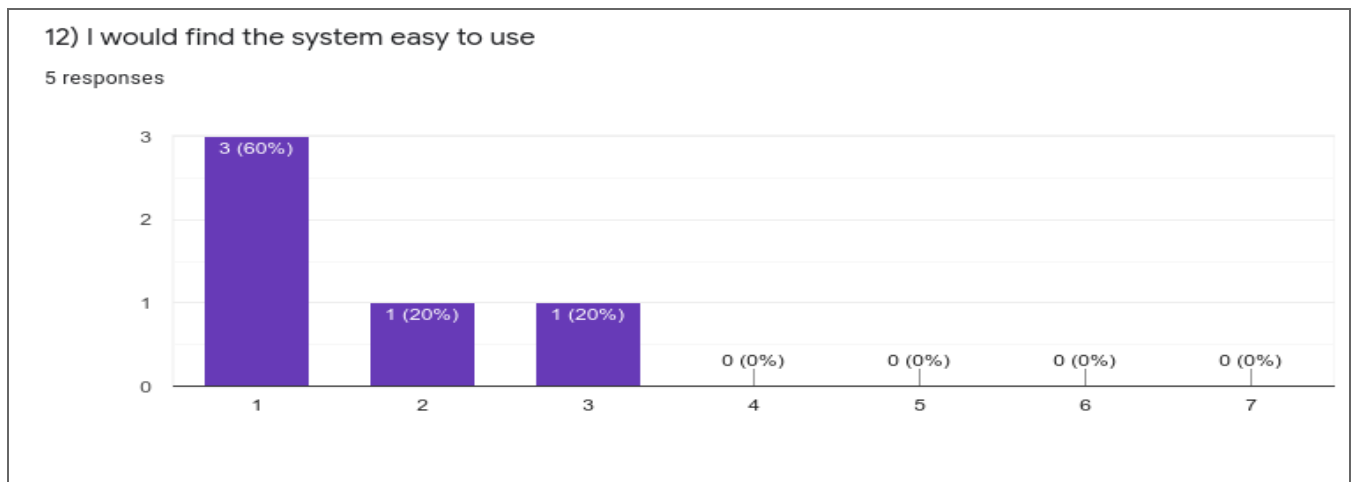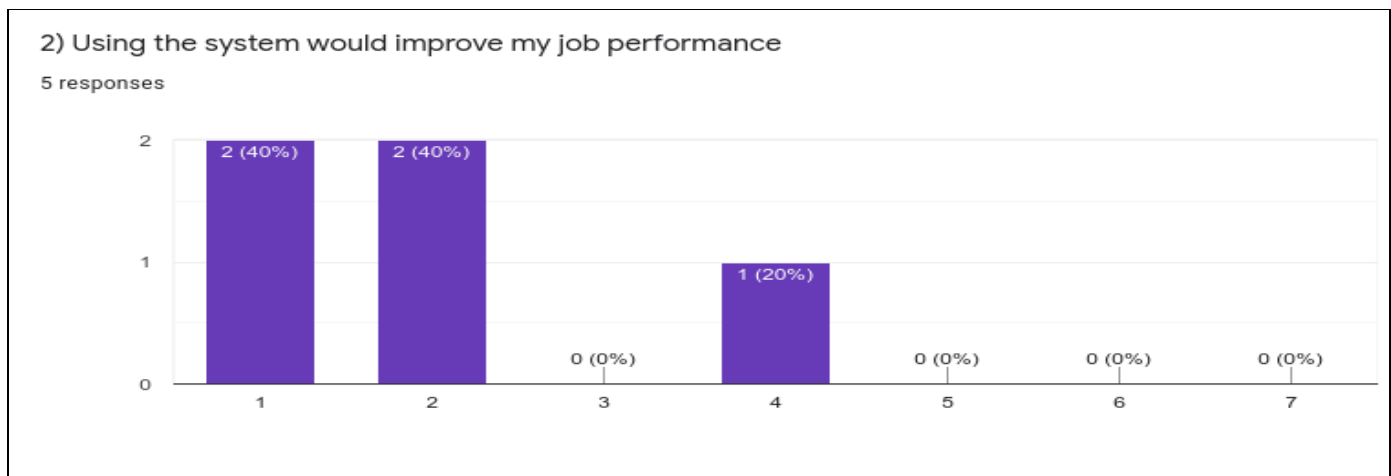

Figure :Number of participants who found the application easy to use
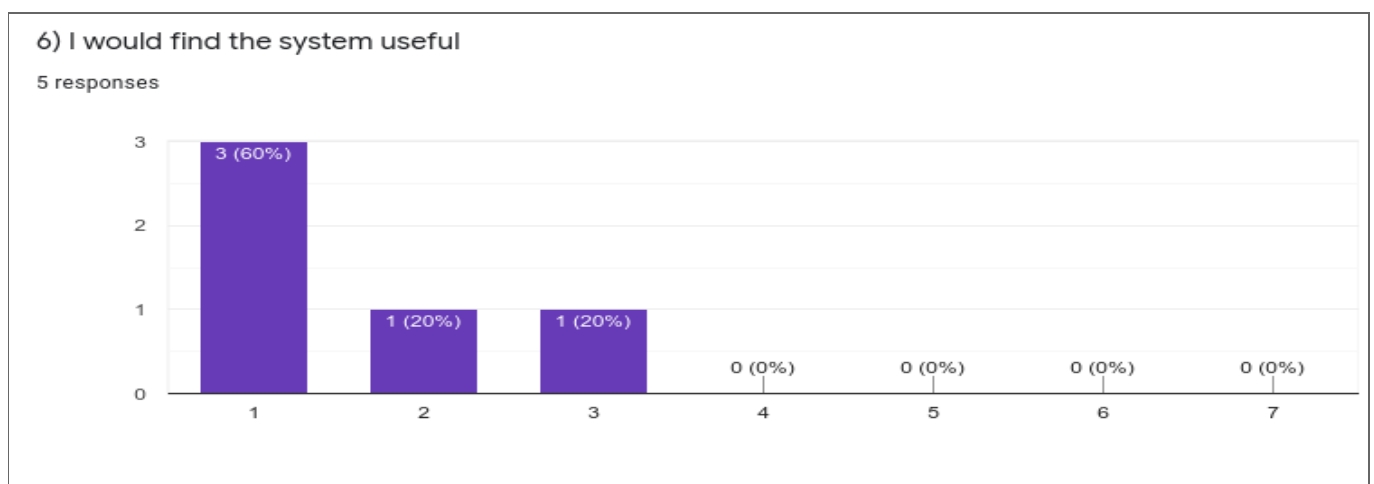
Figure :Number of participants who found the application useful
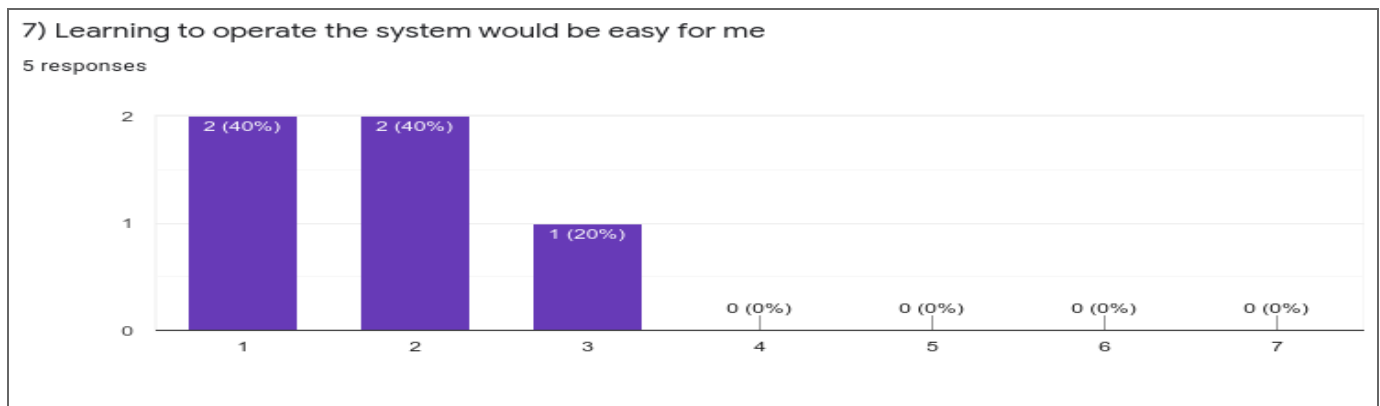
Figure :Number of participants who thought learning to use the application was easy

Figure :Number of participants who thought that the system would improve their job performance

2.4 Explanation of results

2.4.1 *TAM Results:* Figure above shows that 3 participants strongly agree that they found the SAT application easy to use. Looking at Figure above, 3 participants also strongly agree that they find the SAT application useful. Figure above shows that 2 participants strongly agree and 2 others agree that learning to use the SAT application would be easy for them. Figure above shows that 2 participants strongly agree and 2 others agree that using the SAT application would increase their job performance.

Experiment Results: Participants were told to login into Application A and Application B , Then they were to interact with both of these Applications using the same details which were provided and after interacting with the two applications they were requested to fill out the questionnaire respectively. After this task was completed by the participants the results were that the time taken to complete the task for

application A took a long time while the time taken to complete the task in application B took less time than application B.

After that , participants discovered that Application A was DSpace and Application B was UNZA SAT.Responses obtained from participants through the questionnaire showed that UNZA SAT was more effective and easy to use than DSpace.

THE CURRENT PROCESS OF UPLOADING CONTENT TO THE UNZA IR AND ITS WEAKNESSES

Research has reviewed that the current process of uploading IR content is tedious and takes a lot of time and despite the University of Zambia having the repository for a long time, most staff and faculty still cannot get familiar with it, even the staff that is familiar with the process still make a number of errors while uploading content, mistakes such as putting a document in a wrong community, the repository has communities, these communities represent schools and other sectors therefore when uploading content it is important to put content where it belongs for instance a lecturer from the School of Law has to upload content in the school of law community and content from the School of Medicine should be uploaded in the community belonging to the School of Medicine.

The library has quality guidelines that one has to follow when uploading content but most people who upload content tend to make mistakes and do not follow the guidelines, such as putting wrong metadata, the important metadata that should be considered when uploading content in the IR is Author, Title, Subject and Abstract which gives the overview of the material to be uploaded. Other important components that need to be considered are vocabulary control, document quality control and making sure that the links are available if there should be any.

Hence the need to create a tool that is able to do some of these tasks for the user so that the burden is lessened. The Self Archiving Tool is designed to extract content from the material so that a user does not have to do it manually and also the tool will help in automatically doing the vocabulary control, this will cut on some processes and hence make the self archiving or uploading of content to the repository easy and effective.

6 FUTURE WORK

The future work includes synchronizing content with the UNZA Repository and automatic extraction of metadata from the uploaded PDF documents.

6.1. Synchronization:

This synchronization feature shall enable the administrators of the SAT application to synchronize content with the UNZA Repository easily.

6.2. Automatic Extraction:
This feature shall be able to extract the needed metadata from the uploaded PDF documents and automatically populate the form on the SAT application. This feature shall ease the process of uploading documents to the application.

7 CONCLUSION

The introduction of self-archiving among the scholar community revolutionized the way scholars publish their work. The art of Self-Archiving as a whole was an attempt to make the publishing and storing of Scholarly work on repositories easy and to make the scholars' work readily available to the community of their choosing. Despite all these supposed advantages or benefits of self-archiving, repositories still record a low uptake of digital content especially individual deposits which record a negative performance of self-archiving which should not be the case because self-archiving is considered to be an integrated part of the development and practice of IR's. The university of Zambia (UNZA) like many other institutions do practice the art of self-archiving and one might think that because it is a higher learning institution, it would record the highest uptake of digital content to its IR (UNZA repository) but that is not the case, it is imperative to note that the uptake of digital content to the UNZA repository is not really as a result of the practice being difficult. But as a result of the practice not being fully understood by the scholars.

Research undertaken in this research showed that scholars at UNZA had different reasons as to why they do not fully practice self-archiving, reasons varied from the practice being way too complex, practice consumes a lot of time and this brings us to how self-archiving at the UNZA can be improved. One of the ways that self-archiving can be improved is through the development of self-archiving software tools that can make the seemingly tedious and monotonous ways of depositing content to the IR at UNZA effective and easy, which was the sole purpose of this research. The user-friendly Self-Archiving Tool (SAT) will not only improve the way scholars archive their work but also increase the uptake of faculty staff authored pre-prints and post-prints, and Electronic Theses and Dissertations (ETDs) in the IR UNZA. This shows that self-archiving has not lost all hope and value, it just needs improvement and that is what the SAT tool aims at improving, which is the practice of self-archiving at UNZA.

REFERENCES
[1] Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: https://doi.org/10.1145/1188913.1188915.

[2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774

[3] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI:https://doi.org/10.1007/3-540-09237-4.

[4] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY..

APPENDICES

APPENDIX A

QUESTIONNAIRE

This appendix provides information related to the questionnaire design in A.1.

Figure A.1. Showing a screenshot of the online Technology Acceptance Model questionnaire (page 1 of 6)

How many years of experience do you have with using the internet? *

○ Less than 1 year

○ 1-2 years

○ 3-4 years

○ 5 years and above

Section B
Perceived Usefulness

1) Using the system in my job would enable me to accomplish tasks more quickly *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

2) Using the system would improve my job performance *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

Figure A.1. Showing a screenshot of the online Technology Acceptance Model questionnaire (page 2 of 6)

3) Using the system in my job would increase my productivity *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

4) Using the system would enhance my effectiveness on the job *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

5) Using the system would make it easier to do my job *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

6) I would find the system useful *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

Figure A.1. Showing a screenshot of the online Technology Acceptance Model questionnaire (page 3 of 6)

Section C
Perceived Ease of Use

7) Learning to operate the system would be easy for me *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

8) I would find it easy to get the system to do what I want it to do *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

9) My interaction with the system would be clear and understandable *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

Figure A.1. Showing a screenshot of the online Technology Acceptance Model questionnaire (page 4 of 6)

10) I would find the system to be flexible to interact with *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

11) It would be easy for me to become skilful at using the system *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

12) I would find the system easy to use *

	1	2	3	4	5	6	7	
Likely	○	○	○	○	○	○	○	Unlikely

Figure A.1. Showing a screenshot of the online Technology Acceptance Model questionnaire (page 5 of 6)

Figure A.1. Showing a screenshot of the online Technology Acceptance Model questionnaire (page 6 of 6)