

Towards a Web-based Digital Repository: Identification of Needs and Behavior of Users

Wasiwasi J. Mgonzo

The Nelson Mandela African Institution of Science
and Technology
P. O. Box 447, Arusha
Tanzania

Zaipuna O. Yonah

The Nelson Mandela African Institution of Science
and Technology
P. O. Box 447, Arusha
Tanzania

ABSTRACT

Web usage behavior of users have an impact on the development of a web-based application. Web usage analysis is therefore necessary so as to better serve the needs of a web-based application. The process involves analyzing interaction of a user with a web application. Web usage analysis provides useful information that can help improve the service delivery and functioning of a web-based system. The aim of the study reported in this paper is to elicit and study users' needs and requirements towards establishment of an institutional web-based repository for scholarly communication. Questionnaires were used to collect data that reflected web usage behavior and the underlying technical, legal and financial requirements. Analysis results show that there is high usage of the internet as a source of scholarly information and users access their institution websites regularly. Although there is no proper resource management system, the legal requirement for publishing guarantees high research throughput. Results also show that majority of researchers consulted support repository establishment as alternative publishing platform to remove subscription and distribution barriers imposed by journals. In addition, most web usage studies have been system focused. Statistics extracted from web log files and web usage behavior from the users' perspective revealed useful information from which a requirements specification model is developed for the proposed repository system.

General Terms

Requirements, repository, user needs, user-centered approach.

Keywords

Web usage; web usage analysis; web-based application, search engines.

1. INTRODUCTION

An institutional repository is a web-based collection of scholarly materials. It is an institutionally defined collection of scholarly research outputs that provide open or free access to all users within and outside the institution and complies with global standards for interoperability [1]. Institutional repositories have increasingly been recognized as important tools for scholarly communication and are a key source of institutional visibility and a workable tool for knowledge management within an institution.

Within the information environment, a digital repository contains a set of services for collecting and sharing digital objects. They are mostly used as alternative publishing platforms by higher learning institutions and, particularly, more use is observed in research based institutions. As a means of providing wider access to research information, digital repositories form part of the great open access movement worldwide that advocates open access to all

scholarly information. Tanzania has about 28 universities, 19 colleges and 14 centers all providing education at degree, graduate and postgraduate levels [2] but to date only five are implementing institutional repositories.

The primary job behind successful implementation of an institutional repository system is identification of users' needs and requirements. User requirements refer to the features or attributes a product should have or the functions it should be able to perform as seen from the user' perspective. To successfully identify users' needs and requirements, a needs analysis has to be conducted together with eliciting expected system functionalities from the users' perspective, simply because requirements are a sum total of users' needs and expectations.

Requirements elicitation has been defined as the process of seeking, uncovering, acquiring, and elaborating requirements for computer based systems. There are misconceptions that requirements are just captured or collected but in fact, they are elicited. As it employs multiple techniques, approaches and tools for performing them, requirements elicitation is a very complex process. The choice of which technique or approach is appropriate depends on the relative strengths and weaknesses of each technique or approach and the context and situation to be used [3].

In the study reported in this paper, a user centered approach to requirements elicitation is employed. User-centered approach (UCD) is a discipline for collecting and analyzing users' requirements. It is a product development approach that focuses on the end users of a product, and is accomplished by employing user-centered techniques, processes, and methods throughout the production lifecycle [4]. In this case, the users are the *faculty*, *staff* and *students*. To guarantee success, the repository should meet the needs of the users as well as the long term strategic objectives of funders and the institution so that the institution's mission and vision are all met. Exploring user requirements and prioritizing them in the development of the repository and investigating the possible benefits users will have is therefore of paramount importance in the study.

The need for scholars to share their research results more easily and quickly have been a major motive behind establishment of institutional repositories. This has been attributed to a great extent by inability of the traditional publishing system to provide similar services. Despite this wish, scholars from developing countries have not benefited much from such initiatives. The uptake by higher learning institutions of the move has been very slow. However, recent studies show significant progress towards implementation of institutional repositories in Tanzania [5].

In addition, researchers in developing countries like Tanzania are using more informal methods for sharing such

information. These include emails, peer-to-peer mechanisms and informal personal networks. These have proven to have little impact in terms of coverage and efficiency and are not feasible to a growing research institution with a huge volume of scholarly outputs coming every year. To support these less formal methods, an institutional based repository is needed to complement the formal journal publishing system by depositing copies of published works in the repository for easy management, accountability to funders and easy access by users.

Since early 2000s, an increasing number of academic institutions worldwide have established Institutional Repositories to collect and provide open access to their local digital content produced by their faculty, students, and staff. Some repositories in Africa have matured and grown significantly in the past few years [6], with many repositories recording downloads in the millions. Currently Tanzania has 5 institutions that have established institutional repositories having contributed about 3684 records in the global community of scholars. The provision of resources is however not sufficient to improve access. Although awareness of open access remains low in most developing countries, encouraging best use of available resources is the next step towards successful open access repositories [7].

This paper has five main sections. After the introduction in section 1, the rest of the paper is organized as follows. Section 2 presents literature review, Section 3 covers materials and methods, and Section 4 presents the results and discussion and Section 5 carries the conclusion.

2. LITERATURE REVIEW

Developments in information and communication technologies (ICTs) has had an impact on how people collect and share information. The internet has revolutionized these ICTs with greater impact on information exchange than traditional telephone and television combined. It has generally enhanced the collection, storage and distribution of information. These developments have motivated higher learning and research-based institutions worldwide to seize this opportunity towards designing systems that collect, preserve and provide access to their scholarly works for the benefit of other scholars around the world and for academic stewardship.

Studies show that faculty generates enormous amount of information, which is untapped and not available for access [8]. Open access to such research information has been seen to increase research impact and institutional visibility and has greater benefits both economically and technologically through creating innovations. Despite these advantages of open access initiatives, scholarly communication lacks viable means for global dissemination, which reduce the global visibility of many publications from the developing countries [9].

An increasing number of research articles are accessed through the web through various search engines. Through analyzing web log files, useful information can be drawn upon. Various authors have been interested in analyzing how these web usage statistics can be used as predictors for citation impact [10] of using research articles. Others have used similar statistics as indicators of a successful repository [11]. Measuring repository success through web usage data was aimed at discovering usage behavior such as pages requested and number of items most downloaded from web servers.

Reported in [12] is a baseline survey of research materials held in departmental and personal web pages in order to provide a quantitative view of web usage across different subject areas. It was found that these initial data are important and would aid in the initial population of the repositories as they contain useful information that can be used in identifying ready materials and willing scholarly contributors. Success of a repository depends on proper content recruitment plans and is also affected by users' willingness and readiness towards self-archiving their materials in the institutional repository.

Similar studies are reported in [4]. In their studies, authors used web mining and content mining techniques to analyze web server log files using LexiURL and recorded summarized statistics of lists of links (URLs). It was found from the study that information such as who is accessing the system and when, and who could be using the system resources if they are surfing a particular part of a web, could eventually reveal web usage statistics. This information is very useful in determining what materials are more accessed and from which user group.

In [3], while studying visibility of institutional repository contents, authors conducted a study aiming at measuring the accessibility of content from a university repository by search engines and identified that, many users insist on free form internet searching for academic data, instead of academic databases designed for that purpose. The research provides valuable information regarding the need to design effective repositories within the university research environment. Based on the foregoing, it can be argued that analyzing web usage behavior from user perspective has not been thoroughly studied. In Tanzania, for example, great work has focused on determining factors that affect open access adoption and greater efforts have been made to raise awareness [13, 14].

3. MATERIALS AND METHODS

The reported study adopted a survey research design. Online Questionnaires were used as a method of data collection. Online questionnaires were designed using Google docs and were then administered to respondents through emails to 210 (100%) respondents. Data collected were then downloaded from Google docs to Microsoft Excel for analysis. Data were then cleaned for unwanted information. The response rate from the study was reasonable, with 206 (98%) responses. Table 1 shows the distribution of questions (in four main categories) that were of interest in the study.

Table 1: Question design and distribution

Section	Attributes to be identified
Users' characteristics	<ul style="list-style-type: none">• Level of education• Job role
Needs assessment	<ul style="list-style-type: none">• Number of research outputs produced
Technical challenges	<ul style="list-style-type: none">• Preferred publication medium• Preferred forum for sharing research findings• Submission formats of scholarly outputs• Management of scholarly outputs
Legal challenges	<ul style="list-style-type: none">• Legal factors for publishing• Intellectual property rights ownership
Financial challenges	<ul style="list-style-type: none">• Usage of subscription based resources• Access levels of subscribed materials

3.1 Area of study and population

The study was conducted at The Nelson Mandela African Institution of Science and Technology (NM-AIST) in Arusha Tanzania. The population of the study involved faculty members, staff, and graduate students. The choice of the study area is due to the fact that, it is the only research-based African science and technology institution in the country and therefore a large quantity of research outputs is expected from it. This fact calls for the establishment of a web-based repository so as to share these valuable resources with the community around and the global society.

3.2 Data analysis

Both qualitative and quantitative data for the study were analyzed and presented. Quantitative data were codified and entered into Ms. Excel software for analysis. Data analysis, a process through which data is evaluated using analytical and logical reasoning to examine each component of the data, was then carried out using Microsoft excel data analysis tool.

4. RESULTS AND DISCUSSION

This section presents and discusses the findings of the study conducted at NM-AIST. The study findings are presented in the following subsections basing on the four key investigated aspects: *policy issues, management issues, access issues and web usage behavior of users* of an institutional repository.

4.1 Background of respondents

A total of 206 respondents were involved in the study. In describing the characteristics of respondents, level of education and job role were used. It was found that 58 have a PhD level of education (28%), 124 have a masters level (60%), 22 have graduate level (11%) and only 2 (1%) have other levels of education. In terms of job role, the study involved 45 faculty (22%), 23 staff (11%), 138 graduate students (67%) and none (0%) from other categories of job roles. Distribution of the respondents by level of education and job role is summarized in Table 2.

Table 2: Characteristics of Respondents.

Level of Education			Job Role		
Level	Frequency	Percentage	Role	Frequency	Percentage
PhD	58	28%	Faculty	45	22%
Masters	124	60%	Staff	23	11%
Graduate	2	11%	Student	138	67%
Others	2	2%	Other	0	0%
Total	206	100%	Total	206	100%

4.2 Web usage behavior

Adoption and awareness have been identified by various scholars around the world as the main factors that affect open access adoption and usage among institutions and scholars [13, 15]. The other factor that has not received considerable attention is web usage behavior among scholars. To investigate this, we used users' behaviors towards sources of scholarly information, the search engines usage, frequency of accessing their institution's websites and content interest for materials posted in the institution websites. Table 3 shows that majority of respondents (61%) use the internet, 6% use the library, 32% use online journals and 1% prefer other print media. These statistics show that internet usage among researchers' is more ideal for the purpose of this study. With respect to search engine usage, findings show that 50% use Google, 48% use Google scholar, 1% use Bing and 1% preferred use of other search engines. The first two search

engines counts for more usage because of their greater visibility around the world.

Results also reveal that majority of respondents (49%) access the institution website on daily basis, 48% access on weekly basis, 3% access it once a week and none claimed not to have accessed it. In terms of content, 50% are interested in website updates, 25% accessing the e-library resources and about 23% read emails with very few (2%) visiting the website for monitoring students' progress reports. While poor research conditions have been reported to slow down open access usage [13], the results reveals high usage among scholars. Low content recruitment demotivates users because many institutional web pages are frequently out of date or incomplete [16]. Table 3 summarizes the web usage behavior.

Table 3: Web usage behavior of respondents.

Question Item	Attribute	Frequency	Percentage
Major Sources of information	Internet	126	61%
	Library	12	6%
	Online Journals	66	32%
	Other print media	2	1%
	Total	206	100%
Search engines usage	Google	103	50%
	Google scholar	99	48%
	Bing	2	1%
	Other search engines	2	1%
	Total	206	100%
Website usage	Daily	100	48%
	A few times a week	99	49%
	Once a week	7	3%
	Never	0	0%
	Total	206	100%
Content search behavior	Read e-mails	48	23%
	Access the e-library resources	51	25%
	Monitor student progress reports	5	2%
	Read website updates	102	50%
	Total	206	100%

4.3 Needs assessment

The need to develop an institutional repository were derived basing on the number and categories of research outputs. A good number of research outputs is produced at the institution. More results are reported on research articles where most researchers had at least 1 article (96). Following is theses and dissertations, where 49 respondents had at least 1 dissertation/thesis. With research articles 47 have produced more than 2 articles, 14 have more than 3 articles and 15 have more than 4. Low output are observed in teaching and learning materials and patents. Overall, there is high production of research articles as compared to other outputs. These figures can be used as a basis for the establishment of an institutional repository. The results are summarized in Fig. 1.

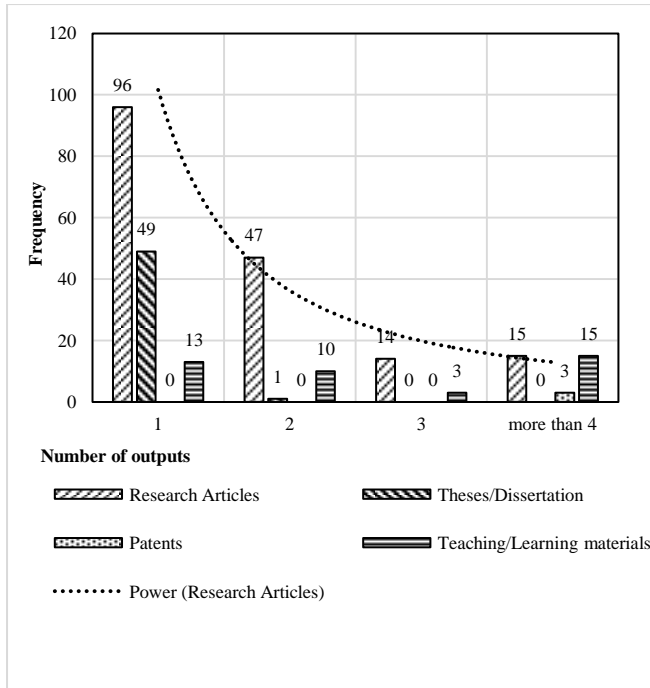


Figure 1. Number of research outputs by category

4.4 Technical challenges

Technical challenges of a web based digital repository were identified by looking at the media used by researchers in publishing their research findings, their local forum for sharing research outputs, submission formats for highly produced scholarly outputs and content management issues. Figure 2 shows that majority are very positive with respect to publishing in a reputable journal (about 92%). A good number preferred publishing through the institutional repository; few optionally prefer personal websites and many did not prefer other communication media.

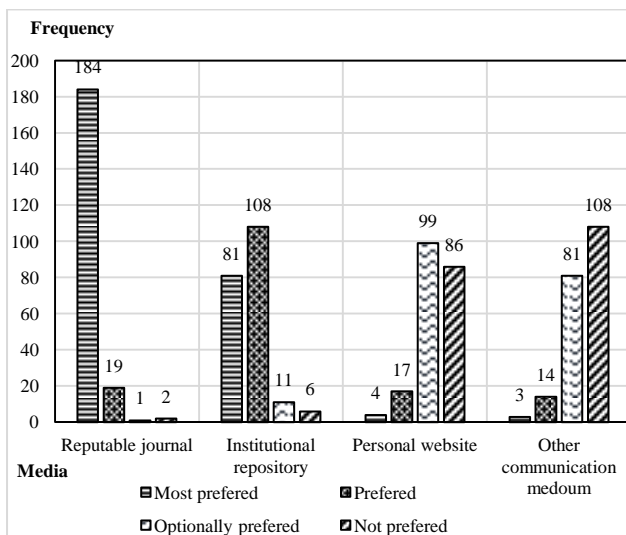


Figure 2: Media preferences among respondents.

Research findings have to be shared and communicated among scholars to be useful and meaningful. Table 4 shows the results of assessing whether there is a forum for sharing and communicating research findings at the institution. It was

found that, 18% uses emails and 77% said there is no such a forum. The submission formats of most scholarly works are in both print and electronic formats (89%). The results about submission formats supports the major purpose of this study. On the other hand, the lack of research findings dissemination and access forum urgently calls for the establishment of an institutional repository.

Table 4: Research forum and content submission format results

Question Item	Attribute	Frequency	Percentage
Institutional research Forum	library	9	4%
	Institution website	2	1%
	E-mails	38	18%
	No forum	157	77%
Total		206	100%
Material Submission format	Print format	14	7%
	Electronic format	7	3%
	Both print and electronic	184	89%
	Other media	1	0%
Total		206	100%

Institutional repositories are designed to collect, preserve and make available the digital collection to all of would be users. If a system does not support any of these functions, then it will not be as successful as the one implementing all functionalities. Authors were interested in finding out challenges encountered in the management of intellectual resources. Figure 3 shows that, a significant number of respondents agree that there is no preservation system (115), majority reported about lack of resource sharing policy (144) and access mechanism (155) and access mechanism (57).

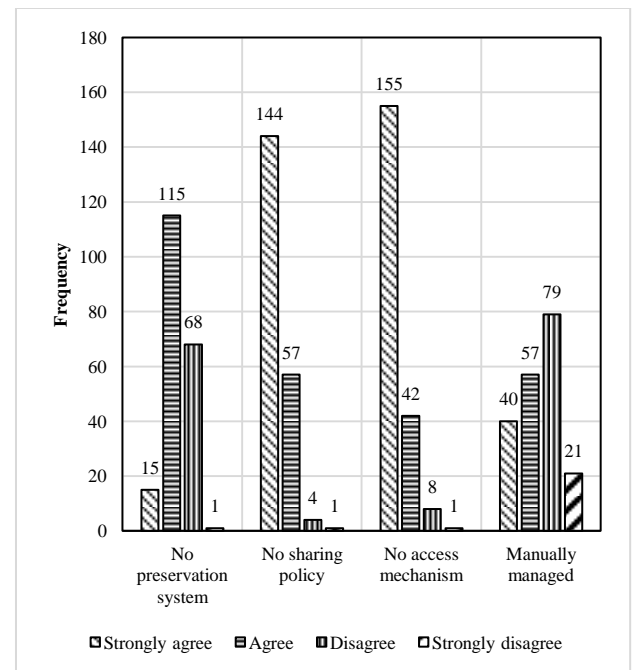


Figure 3: Management of scholarly outputs.

4.5 Legal challenges

Motivational factors towards publishing and factors regarding copyright issues were studied in order to assess the legal requirements of an institutional repository. The assessment results are summarized in Fig. 4. Further, it is confirmed in [7] that, various policies have been proposed and adopted at various organizations and that these policies have shown to have impact on repository content development and accumulation. Figure 4 shows that majority of researchers

strongly agree that publishing is mainly based on two factors: graduation and promotion. Subsidized publication costs have also been shown to motivate publishers while personal motivation received low scores. While copyright issues are cited to have a great impact on repository content recruitment, majority of respondents (84%) indicated that the institution owns property rights for all institution funded research. The results are as shown in Fig. 5.

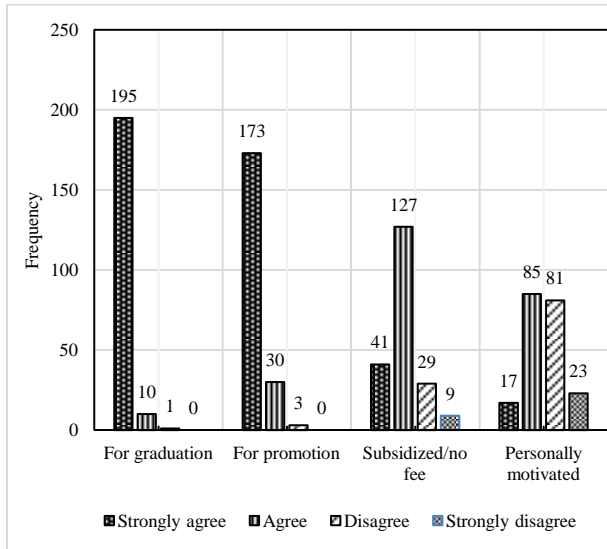


Figure 4: Factors that motivate publishing among NM-AIST scholars.

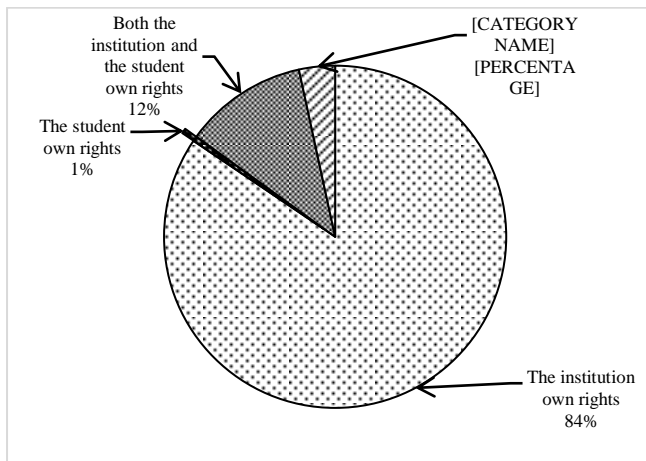


Figure 5: Intellectual property rights management

4.6 Financial requirements

There are subscription based constraints on accessibility [11, 18], and these have been used by most commercial publishing companies to restrict access to scholarly information. Table 5 shows that 60% of respondents have used subscription-based materials while 40% have not used. While great efforts have been made by the institution to subscribe to a number of journals, the use of materials from such journals is highly restricted to members of the institution only (48%) and some respondents reported to have no information about these resources (24%). This could be due to the access barriers imposed by journals [16] and authentication process complexities involved in the access process.

Table 5: Usage characteristics of subscription-based materials

Question Item	Attribute	Frequency	Percentage
Use of subscription based journals articles	Yes	124	60%
	No	82	40%
Total		206	100%
Access level of Institutional subscribed e-journals	Not Accessible	45	22%
	Restricted to NM-AIST only	98	48%
	Openly accessible	13	6%
	No information	50	24%
Total		206	100%

5. CONCLUSION

The study reported in this paper was conducted to elicit requirements for the development of a web-based digital repository for scholarly communication through studying users' web-usage behavior. The study was based on the fact that web usage behavior has an impact on successful implementation of any repository system. Based on the findings from the survey, the following can be noted about the NM-AIST community and in general:

- Majority of respondents are participating in research activities in one way or another.
- There is high internet usage and website access frequencies with majority using free form search to locate materials online and are interested in any new updates in the web pages.
- There is a considerable high throughput of research articles and theses / dissertations at the institution.
- Establishment of an institutional repository has a strong user support as most of them prefer using it as an alternative publishing platform and have reported on the absence of a dedicated platform for sharing their research findings. Again, the content submission format employed by the institution supports the required data format for the repository system.
- There is lack of a proper resource management system and available policies do not support it. These policies could include resource sharing policies, access mechanism defined policies and preservation policies.
- On one hand, there is a legal requirement for publishing which guarantees content recruitment. On the other hand, copyright and other intellectual property rights rests in the custody of the institution. Use and distribution of any scholarly output can be decided at institutional level.
- While there has been efforts to subscribe to journals by the institution, the approach does not remove any access barriers imposed by journals simply because traditional publishing companies still restrict fair use like distribution to any other user who uses the materials for fair dealing as is allowed by the open access movement under creative commons rights management policies.

Generally requirements are description of the system services and constraints that are generated during the requirements engineering process [19]. The end result of any requirements engineering process is a requirements specifications. The

result of the study concludes by presenting the requirements specifications for the proposed repository system as shown in Fig.6.

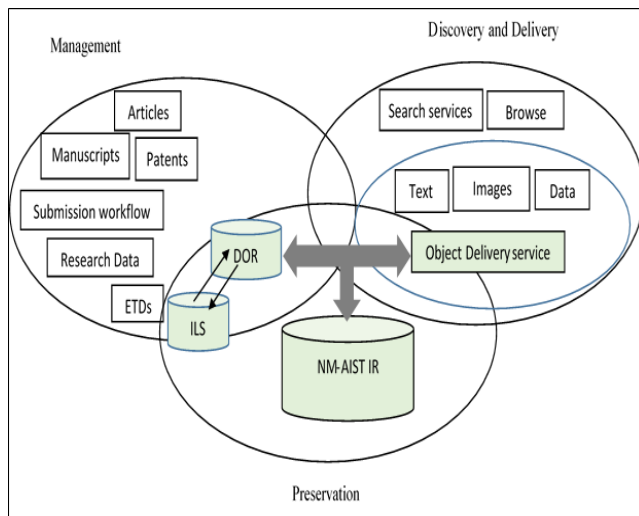


Figure 6: Requirements specifications for NM-AIST repository

KEY: ILS - Integrated Library Services, DOR - Digital Objects Registry, ETDs - Electronic Theses and Dissertations, IR - Institutional Repository.

The requirements engineering and specification process identified the services customers require from the system and constraints under which it is to be developed and operated. These services are generally referred to as requirements (in technical terms) and are categorized as:

- a) **User requirements:** Statements in natural language plus diagrams of the services that the system provides and its operational constraints. The user requirements have been identified and discussed throughout in the results section of this paper.
- b) **System requirements:** A structured document with detailed description of the system services, which defines what should be implemented. System requirements can further be organized as:
 - i. **Functional requirements:** Statements of services the system should provide, how the system should react to particular inputs and how the system should behave in particular situations.
 - ii. **Non-functional requirements:** Constraints on the services or functions offered by the system such as timing constraints, constraints on the development process, standards, etc. [20].

System requirements which depicts high level system functionality to be provided to satisfy user needs are derived from users' expectation and have been identified from the study as:

Functional requirements:

- The system should be able to provide storage functionality (preservation) to all the scholarly outputs of the institution

- The system should provide an online forum where users can access, view and share research materials.
- The system should provide free access to all the contents stored in the system.
- The system must be able to provide statistics of items recorded in the repository together with the usage statistics for management purposes.
- Authors should be able to create accounts so that work submitted to the system is monitored and controlled for quality.

Non-functional requirements:

- The user interface for repository system shall be implemented as simple as possible to avoid complex navigation system.
- The security mechanisms have to be implemented to make sure the resources in the repository are protected from access by intruders in the storage layer.
- As a web-based application, the network should be reliable to provide 24/7 access to the repository.

The requirements identified in this study can be used as the basis for the design and implementation phase in the proposed digital repository system development process. The challenges of implementing the system include adoption of relevant global standards for interoperability, choice of suitable software, integrating with other information management systems, which includes redefining the submission workflow to incorporate new functionalities.

6. ACKNOWLEDGEMENT

The authors acknowledge the Nelson Mandela African Institution of Science and Technology (NM-AIST) specifically the school of Computation and Communication Science and Engineering (CoCSE) for the support that made this work a success.

7. REFERENCES

- [1] M. Ware. "Institutional repositories and scholarly publishing". *Learned Publishing*, Vol. 17, pp.115-124, 2004. [Accessed: August 15, 2014].
- [2] TCU. "Register of Universities". Internet: <http://www.tcu.go.tz/index.php/register-of-universities> [Accessed: August 15, 2014].
- [3] V. Kiva and M. Weideman. "Visibility of institutional repository content", in *Annual Conference on World Wide web Applications*, 2013.
- [4] A. Zuccala, M. Thelwall, C. Oppenheim et al. "Web intelligence analyses of digital libraries: A case study of the National electronic Library for Health (NeLH)". *Journal of Documentation*, Vol. 63, pp.558-589, 2007. [Accessed: August 25, 2014].
- [5] W. J. Mgonzo and Z. O. Yonah. "A Review of Open Access Publication in Tanzania". Unpublished.
- [6] RWR. "Ranking Web of Repositories". Internet: <http://repositories.webometrics.info/en/Africa/Tanzania%2C%20United%20Republic%20of> , July 30, 2014 [August 18, 2014].
- [7] M.A. Islam and R. Akter "Institutional Repositories and Open Access Initiatives in Bangladesh: A New Paradigm

- of Scholarly Communication". *Liber Quarterly*, Vol 23, pp.3-24, 2013. [accessed: August 25, 2014]
- [8] P. Chand, T. Murthy, K. Prakash and U. Gohel. "Institutional Repositories, Open Access Movement and OAI-PMH Complaint Software," 2nd *Convention PLANNE-2004*. Manipur Uni, Imphal. 2004, pp. 4-5.
- [9] I. J. Ezema. "Building open access institutional repositories for global visibility of Nigerian scholarly publication", *Library Review*, Vol. 60, no.6, pp. 473-485, 2013. [accessed: August 25, 2014]
- [10] T. Brody, S. Harnad and L.Carr. "Earlier web usage statistics as predictors of later citation impact." *Journal of the American Society for Information Science and Technology*, Vol. 57, pp. 1060-1072, 2006. [Accessed: August 25, 2014].
- [11] M. Westell. "Institutional repositories: proposed indicators of success". *Library hi tech*, Vol. 24, pp.211-226, 2006. [Accessed: August 15, 2014].
- [12] T. Andrew. "Trends in self-posting of research material online by academic staff", *Ariadne*, Vol. 37, 2003.[Accessed: August 18,2014]
- [13] F.W. Dulle, M.Minish-Majanja and L. Cloete. "Factors influencing the adoption of open access scholarly communication in Tanzanian public universities," *World Library and Information Congress: 76th IFLA General Conference and Assembly*. Gothenburg, Sweden.2010.
- [14] F. Dulle and M. Minish-majanja, "Researchers' perspectives on open access scholarly communication in Tanzanian public universities," *South African Journal of Information Management* , Vol. 11, no. 4, pp. 1-14, 2009. [Accessed August 2014].
- [15] A. Moller. "The case for open access publishing, with special reference to open access journals and their prospects in South Africa" Masters' thesis, University of Western Cape, South Africa, 2006.
- [16] A. Swan and L. Carr L. "Institutions, their repositories and the web". *Serials review*, Vol.34, pp. 31-35, 2008. [Accessed: August 25, 2014]
- [17] Y.Gargouri, C.Hajjem C, V.Larivière et al. "Self-selected or mandated, open access increases citation impact for higher quality research". *PloS one* 5, e13636, 2010.[Accessed: August 25, 2014]
- [18] P.Suber. (2005). *Open access overview*. [Online]. Available: [http://cyber.law.harvard.edu/hoap/Open_Access_\(the_book\)](http://cyber.law.harvard.edu/hoap/Open_Access_(the_book)) [July 18, 2014].
- [19] E.Hull, J.Keneth, D.Jeremy. (2005). *Requirements Engineering*. [Online] Available: <http://link.springer.com/book/10.1007/b138335> [Accessed: September 7, 2014].
- [20] I. Sommerville. *Software Engineering*. 7th Edition. Addison Wesley: 2004, chapter 6.