Vol. 4 · July 2014
Print ISSN 2244-1824 · Online ISSN 2244-1816
doi: http://dx.doi.org/10.7719/irj.v4i1.303

JPAIR Institutional Research is produced
by PAIR, an ISO 9001:2008 QMS certified
by AJA Registrars, Inc.

Systems Development for Records Archiving and Digital Documents Repository: A Case Study

MICHAEL P. GAMBA
http://orcid.org 0000-0003-1108-3094
gamba.michael@gmail.com
Sorsogon State College
Sorsogon City

MAGDALENA M. OCBIAN
http://orcid.org 0000-0002-2421-2548
mmocbian@gmail.com
Sorsogon State College
Sorsogon City, Philippines

MARYJEAN N. GAMBA
http://orcid.org 0000-0002-4033-3294
jeangamba27@gmail.com
Sorsogon State College
Sorsogon City, Philippines

ABSTRACT

The presence of vast communication and information nowadays necessitates the need for a system to readily access and transfer data. The study aimed to develop records archiving and document repository to overcome the barrier of server-client method of deploying the documents from one place to another and easier data access to its stakeholders. The Sorsogon State College in the Philippines has four campuses which are located strategically in four municipalities of Sorsogon province. Their distance from each other sometimes causes a problem particularly

along communication and real time updates. General User Interface (GUI) of the application has been built on top of all web-enabled computers and even to the mobile devices, so it requires only installed web browsers to render the GUI onto their devices regardless of its platform and specifications. The installation of web-based archiving and repository on its main campus enables satellite campuses to connect to the college private server in a cost-effective manner through virtual private network that connects on top of the Internet service provider. This study overcomes the vulnerability of security by means of allowing user credentials to login at the private server using a 1024 bit Rivest-Shamir-Adleman (RSA) private/public key exchange and 256-bit Advance Encryption System (AES) encryption through its virtual private network. Contents of the uploaded files were being encrypted at 128-bit to prevent intranet users from sneaking the file contents.

Keywords **-** ICT, document archive, web, GUI, developmental research, Sorsogon City, Philippines,

INTRODUCTION

Records archiving already existed since the ancient times. In every institution, the records support the authenticity of the day-to-day transactions. These records are all information-based materials which serve as references to transactions or activities. Regardless of their format, they are being reproduced and stored for future reference of the institution as it may serve as valuable data for every office or institution.

Institutions' personnel were able to preserve their documents over a long period by storing such pertinent documents in a simple storage room. The composition of these paper-based information/document materials is being degraded by the environmental factors like thermal variations and humidity of the place. Thus, to preserve these materials for a longer period, institution might need an environment that will eventually prevent external factors from affecting the quality of these paper-based documents.

In an academic institution, manuscript records and other library materials are sources of information that contains definite contribution or facts to a certain field. These information are being accessed by students for various purposes from time to time. The mitigating procedure of universities and colleges to maintain the usability of these documents is to store its original copy in a more secured room

where its material and form will be preserved. However, not all information in a university or college is for public domain user. Some are being kept intrinsically for personal use only in separate storage room for limited access. Other users are not given equal chances in accessing important documents.

The National Archive of the Philippines created by Republic Act No. 9470 imposes to move all records or documents with archival value from all branches of the government to a permanent repository. However, this does not only cover the public offices but also the private organizations where documents and other records have bearing to the culture of the country or an essential contribution to the policy makers. This Act also covers the public documents of any government agency that bears a public transparency from its everyday activities (Republic Act No. 9470).

In an agency or institution that demands a dynamic environment, sharing and collaboration of data play a great role to every stakeholder. The ease of accessibility and availability of these data becomes a primary requirement. Thus, records archiving and repositories must be established at a certain place in an institution where every stakeholder can have a direct access to its data in accordance to the possible limitations that it might offer due to confidentiality of some data. This solution might lessen backlogs and achieve greater personnel productivity.

The Sorsogon State College is the only state college in the province of Sorsogon, Philippines. It has four campuses located strategically in the four municipalities of the province to cater the clients in accordance with its vision, mission and goals. However, because of the distance from each campus, retrieval of records and information becomes a problem which lessens the efficiency of some personnel and the productivity of the College, in general.

With this study, the college archives and records repositories will be connected to every campus at one place for a centralized access, for faster data transfer and updates allowing a continuous flow of data anytime of the working day. This strategy will not only lessen the travel cost but will likewise, increase the data access rate to documents for widest dissemination at real time to different campuses. This will eventually lead to greater productivity of every personnel. This study also takes into account AACCUP's recommendations on Administration Area during the Level II Actual Survey last November 2008 that documents should be placed in the storage room and be convert into digital form for easier access of data.

Thus, to extend the usability of such materials and facilitate an easier retrieval, dissemination, exchange and/or storage of records and documents among or between departments or campuses, a system was developed that will convert them into a digital image for an intranet/Internet distribution.

O'reilly (2005) coined the Web 2.0 platform for all web-based software (either intranet or Internet distribution) that allow the users' participation in the contribution of metadata on every site that caters a file hosting service. As time goes by, web repositories contain almost all types of data that can be downloaded starting from documents up to multimedia files resulting to a greater availability and easier access to online resources for every mobile device that was connected to the Internet and intranet distribution.

Shiman et al. (2001) devised a documentary repository in such a way that all employees must retrieve their documents from a central server. To validate the authenticity of the users or editors, each has the authocodes or document tags as embedded in the file after editing. Document servers control the version numbers of the files to indicate the number of revisions that have been undertaken. At some point, users or authors have the option to enter their information which may become part of the document during its distribution process.

Kohl et al. (2013) revealed in their study that patients' records in the health services will be more useful for different organizations or groups if these had been converted into digitized format. The digital records allow personnel of health care to share data for clinical surveys or trials. They are also relieved from routinary physical handling of data for indexing and filing through time. Records conversion requires a standard procedure that complies protocol and integrity of the data. If the converted data follow the standardization on records conversion, these outputs are constantly monitored to maintain its quality. Any alteration will make each and every digitized information to be void especially if it is to be shared to other groups of health professionals.

Gomes, Miranda and Costa (2011) describe the importance of web archiving to represent the past and the future of Internet or web environment. They emphasized that archiving web pages brings great importance to differentiate the improvement of web up to the modern societies. These differences allow funders and investors on what development they could harness in terms of web environment. The study revealed that most of the countries that started the web archiving were mostly initiated from the developed countries. The archived and collected data were tremendously large composed of different file formats that had been used to store and transmit over the Internet. The collected data have

been analyzed using metrics; such volume of data gathered based on geographical locations, data file formats, and people involved in the production of the web pages. Their survey showed that despite the initiatives of the developed countries and large volume of data collected, still these initiatives were considered scarce.

In the academic setting, Buchanan et al. (2012) initiated a procedure for archiving the records of students' activities and campus developments through collecting, arranging, and describing records from one another. Their procedures revealed that only a small portion of the archived documents represented the students' activities. This only shows the minimal development of the social life of students during their college life as presented in their case study regarding records archiving.

The study of Kim (2011) allowed the participation of the institutional stakeholders for archiving. He motivated them to store their materials to become widely accessible over the university. This initiative benefited every stakeholder of Open Access to their institutional repositories, improved digital preservation, and copyright management.

Likewise, Krishnamurthy and Kemparaju (2011) reported that 20 universities in India that have been accessed had unique contents. Hence, the purpose of their study was to identify the commonality of the repositories in these universities to allow an interconnection and sharing of data in the academe. The data collected for the study were based on content type, metadata and characteristics of each file.

The study on approaches of self-archiving at the Harvard University was proven to be successful for sustaining the faculty. Integration of this method allows the stakeholders the familiarity of the current outputs of every faculty. The availability of these data especially on academic authoring enables them to be more comprehensive about the institutional record or view of the published works with other faculty (Giesecke, 2011).

Nadeem (2010) presents a multi-user desktop environment for archiving documents of the institutions including the students' statistical data and staff documents being used in their day-to-day activities. These data are being converted into digital format and being stamped with security barcode for easier tracking and retrieval of the documents from their system. Other documents like forms are converted into fillable documents by using Optical Character Recognition (OCR). The common files that were archived in their institution were usually documents, letters, and certificates. These are automatically saved in their database after they had been processed in a high-speed multi-page

scanning system. This procedure reduces the stress of their personnel in physical handling and storing of the hard copy of the files which result to a large volume of documents being piled up in their repository area. The integrated security barcode of each document carries the unique signature of the file as a form of authenticity.

Further, Pennock and Kelly (2006) did a research project for archiving records from different respondents on the Internet as their sources. Researchers foresee that the enthusiasm of users to upload their documents serves as their online storage where they can download it anywhere and anytime. However, despite the huge amount of data collected that could be valuable to the users who uploaded them, these data are still cannot be considered as authentic records or documents which can be used as reference for legal undertakings.

The presence of digital documents repository and records archiving in an organization enables every personnel to have filing collaboration and elicit knowledge through sharable files from documents to multimedia. Their internal organizational filing enables files to be listed at the top which are frequently retrieved and accessed by personnel. However, since users have their freedom to store files on their own in the internal systems; it shows that the arrangement of files was in accordance with their personal value (Wu & Gordon, 2004).

In addition, Askhoj (2011) did a study on archiving in a cloud computing environment simplify the document lifecycle and records transfer. It was highlighted that storage layers on cloud computing environment provide a more secure environment for data and its backup. The cloud environment used is a server-client in nature but encapsulated on machine, but located outside the current geographical location of its data source. This method of data transfer is more reliable if the data source originates from those places where stability of the environment is a great issue such as earthquake, tsunami, typhoons, and hurricanes. They present the OAIS reference model for document archiving wherein application has been built on top of the operating system layer. Each module of the application has been designed to be a shared process so other users can open the application at the same time without bothering their possible sharing violations.

The study of El-Bakry and Mohammed (2009) justified that efficient document archiving requires the documents' file size to be smaller for data transport. Their conclusion was based on their designed fast neural network to measure the rate of data transfer for various file formats. Results of the study showed that portable document file (PDF) format is the best file format to transfer or to be accessed by the remote users. The results have been verified using the MATLAB simulations

on comparing the theoretical computations between the conventional neural networks and fast neural networks.

Esser (2013) made the data archiving by extracting the information from the documents that had been processed by scanning and converting them into digital format. Most of the documents that have been tested had a definite format or common fields in the contents. These documents can be classified easily in accordance with their purpose because of the common fields in every document. As a result, the indexing and organization of the documents were made easier. This automatic configuration will be applicable only if all documents have a definite format, where small office and home office cannot implement this type of configurations due to various formats that they employ in their day-to-day documents.

The study of Yoo (2014) proposed the use of cloud-based storage system for managing document repositories. The emerging technologies such as mobile devices can easily acquire data or documents because of high availability of Internet connections through wireless-based technology. Most mobile devices have been equipped with basic word processing by porting free and commercialized text processing software. He emphasized that a well-managed document repositories should not allow the end-users to specify the dedicated folders to store their documents and other related data. The storage systems will organized and index the files that have been submitted for a better searching capability.

On the other hand, Seadle (2012) discussed the archiving in the network environment, both intranet and Internet way of transferring data over method. His study was concerned with the integrity and authenticity of the transmitted data. The user can receive and send data but could not validate its integrity. Moreover, the authenticity becomes more important since there is a situation that digital data can now be honored as an exhibit in the court of law. He emphasized that integrity of the data can easily be checked with rigorous mathematical algorithms with respect to its original file. The procedures on checking the file authenticity remains untested due to lack of protocol standardization.

The study is related to the studies presented in its purpose of preserving valuable documents in a different field by means of their repository and allowing them to be accessed by stakeholders to their public servers anytime and anywhere. However, some of the cited studies differed with the current study in terms of accessibility of the repository server. All data stored in the repository can be accessed only by those employees who have the proper credentials and prevent an anonymous access to the documents stored. The repository servers can only be accessed through the software gateway to validate its legitimate connection;

thus, preventing vulnerabilities on its security protocol. The rate of transfer of documents from server to clients depends on file size and format. Moreover, most of the deployed archiving software presented in the studies were server-client in nature which is another form of security for their deployment.

FRAMEWORK

This study utilized the simplified approach on software development for web-based application to make the development faster. The paradigm as shown in Figure 1 had been divided into 6 subparts, namely: environmental scanning; data collection; application specifications and design; application development; testing and integration; and deployment and population of data on the application itself. The arrows indicate the input and output direction for the whole duration of the process development.

Figure 1. Conceptual Paradigm

A minimal amount of time is required to be spent to the end-user's data collection and environmental scanning since its application specifications and design were based on the project initiation form that was filled out by the end-user. Its application development is based on the incremental prototyping technique where application is being divided into a module. The most important modules are being developed first and followed later by the least important ones. Testing is a situation where actual data from the Records Office are being scanned and uploaded to the application itself; integration corresponds to the situation where application is being tested to work with the other pc-based web-browser such as

chrome, Firefox and Internet explorer. The robustness of the mobile devices the layout of the application itself can be tested through their internal web-browser, opera and safari.

Furthermore, the last part of the paradigm is the transferring and changing of its configuration. These were done to adapt the new network environment to accept simultaneous data uploads and access.

OBJECTIVE OF THE STUDY

This research project primarily aims to develop a system for intranet distribution of digital records archiving and document repository for the Sorsogon State College. Specifically, this aims to develop a web-based application for Records Archiving and Digital Documents Repository for virtual private network distribution.

METHODOLOGY

This study is a developmental research since it developed a system for organizing digital documents being archived at the storage room of the Records Office of the Sorsogon State College. Documents are categorized according to transaction types or purposes. However, these document types can be set later on by the end-user to allow flexibility in their categorization of records. The stored physical documents will be converted in digital format combined with some text parameters for classification and searching.

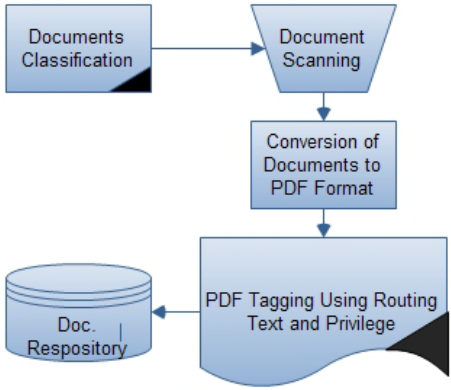

Figure 2. Process of records archiving to establish a document repository

The application specification and design were based on the project initiation form filled up by the Records Office includes the transaction's name and details of its operations. Also, the office manual serves as the basis in the design of its security protocol since they are mandated in accordance with the provisions stated in the national archive law.

The developed system used the web-based format to allow the interaction of other users that will be connected through virtual private network. Basically, it utilized the PHP language for server side script, HTML language for the presentation layer, and MySQL database for internal storage. Optionally, the jQuery framework was only utilized for presentation abstraction to simplify some of the basic processes. The output web application will allow every user to access the repository as long as there is a web browser installed in their mobile device or computer. The accessibility of the data from the document repository server can be retrieved using web-browser only. However, to read or download the data will require specific privilege for every documents available in the repository.

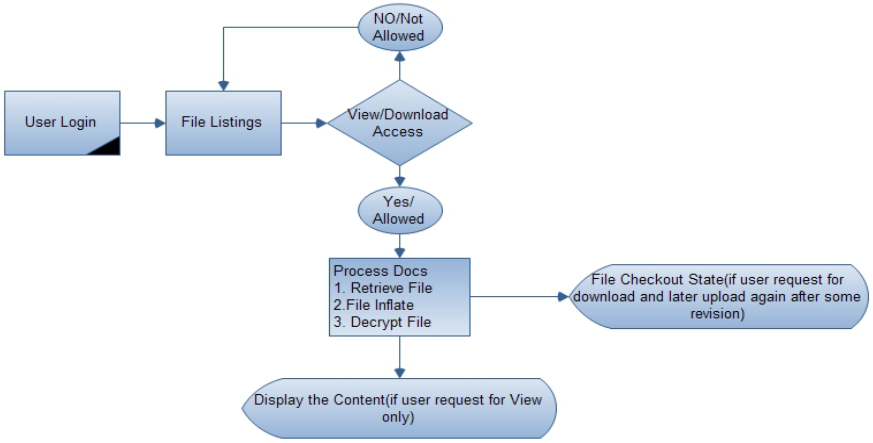

Figure 3. Process flow on how the user can access the data from the file server

The web application and its repository will be installed at the Sorsogon State College-Sorsogon City Campus because of the availability of documents. Thus, other campuses can access the repository through the web application interface using Virtual Private Network (VPN) connections that will be installed to their

machines. The software that was used for VPN connection will be a choice of Hamachi and TeamViewer applications. Both applications are configured to the server to allow a simultaneous connection from legitimate employees of the same institution.

RESULTS AND DISCUSSION

A. Archiving

Data Contents Security. To archive the common documents hard copies that can be found at the Records Office, it will be scanned into digital format preferably the Portable Digital Format (PDF) file. Each file has its own short description that will be manually encoded by the personnel of the office as to classification and searching.

As revealed by El-Bakry and Mohammed (2009), pdf files were faster and efficient for transferring over the network. The consistency of the file contents and embedded attributes and physical configuration of every pdf file being sent over the network makes it reliable. However, not only a simple pdf file can be stored by the archiving software but also the other 18 file formats such as image files, document files, sound files, and others.

During uploads, the file will be deflated in accordance with the security measures which are built-in inside the archiving algorithms, which is good for faster transmission from user to the file server since it is in a smaller file sizes. In this manner, all files will be treated as one format, which will prevent other users from sneaking out the contents of the file. These files will again be inflated if the registered user is allowed to read, download and checkout the file for further changes.

Table 1 shows that after the file uploads, the contents of the file will be encrypted automatically using the 128-bit AES. It can be decrypted only through its associated key-file that is being automatically generated during the file upload.

Table 1. File sample data contents

Before Upload, data contents	After uploads, data contents
204.93.240.0/24 204.93.177.0/24	À¹.O“ñFy¸¡Ùè!WN)ÓÁá¨{ÖJ¦fß%Ã”²AÁ!zò×°‡væÁé…
199.27.128.0/21173.245.48.0/20	$ÚeùO:ÞHEÊý{ËIE^°#á““íºþ‰ø.C.ÇÞË©òŒØkŠ¸A Z%D²igNß8ºÒÇù‡UÀ06ôŽ%žÙÚu¬ÍA¬Rg¨+AŒ?
103.21.244.0/22103.22.200.0/22	ÿË/é-ñ€e]e.›—%FOÓ
103.31.4.0/22,141.101.64.0/18	”º»p?sÍ9ysV´v`Ù?–ÏÓ,ûºüÌ81ÞRíâý:üG½¹Ñ±’.‰‰% n“.².$©%ÏÞ»©xh6¿E¦º¯:3^ó?Juª NÊ¯´
108.162.192.0/18,190.93.240.0/20	¦*œ%a¿l‹x}õ–UÆÂ¿&:ÃrÍœxÅ<YÍ‘¯KÞÔ`mý-ž+mÖ¡¬ e§èq¼”WGsvNHÁÆA¶œ*s„psD¹29l£Ó<’í?Í¸žÆÇ:Ë
188.114.96.0/20,197.234.240.0/22	àT[±ùEÉ„#Ù?ûjÛñ;ÍkC2ðjûÉ1 ’¹]’ÿÍ.áò:SàÌSÓMÒ•Æ
198.41.128.0/17	€¼èÔ⊠ÄSOá$µJ¤·GZ4êCŸõœáÅ¬h“>Ó°€·H°>jo„¥ ÷ñ…6÷îr{ÀçEŠ=@[ðz¤×þDÝTŸ5´2rDìKà+I-p?¯~Bh OU‰¤§Ñ⊠dW;aèmKZ;ß¯Ã3…Ö´ðò”P

The file contents were being scrambled if higher form of attacks to the file server will happen. However, if there is a normal operation where all users must login with proper credentials and accessing the files with proper privileges, these scrambled contents will be decrypted using the associated key file. The decrypted file can now be viewed or even downloaded by the user using the web browser. The response on issue of integrity of the files that can be downloaded from our server had been addressed by using encryption with mutating algorithms over the time to make sure that every file is originally derived from the repository. This method will not allow black hole mechanism since file decryption process can be done only inside the file server machine. As a result, other sites or domains that will host the same file will not be able to render the actual contents of it since the key file only resides in the same server. Thus, similar attack like man-in-the-middle (MITM) can capture file during file uploads and vulnerable for file tampering. Thus, before file upload, a key file is being generated first to secure the original properties of the file such as its size and author/s, as the two data arrives the uploaded file can be checked easily if it was tampered or not by cross-checking with the associated key file during file uploads.

User Logins. This module has been integrated to the application to prevent anonymous users that may randomly browse the files in the repository. The content of the repository were documents from the day-to-day transactions of the Sorsogon State College campus. The confidentiality of some files, the reviewer and the administrator allow only some of the key personnel to view, edit, or

even download the files for further revisions and later uploaded to the server for updates. User logins are composed generally of username and passwords that were hashed through a customized encryption algorithm to prevent brute-force attack or username and password guessing in an automated process using some known cryptography. Users with definite privileges were expected to view, read, download and delete files within the repository through archiving software interface. However, their privilege varies from one credential to another or depends upon the permissions set by the administrator or reviewer during the file uploads.

Application Integration. Most of the software development related to archiving and repository were server-client in nature; however, with the maturity of the technology regarding network infrastructure either intranet or Internet purposes, applications were built on top of every operating system that is meant to be a cross-platform deployment. The archiving interface can utilize every browser found in the client system to render its general user interface; thereby the issues on software compatibility are eliminated. As tested on available web browsers such as Firefox, Chrome, and Internet Explorer, only the latter creates

some minor issues on its layout rendering due to customized framework of the software company. The issues had been fixed with the help of the some experts through online forums of the software company. The layout of its general user interface has been optimized to be rendered easily on IOS and android-based mobile devices. Its interface had been tested on these devices that have the screen range of 8.0 to 10.1 inches such as tablets.

The shared modules have been ported to mobile devices since these equipment were the emerging devices nowadays. Most of all, people will be more comfortable in accessing data because of the inclusion of the basic word processing and rendering common file formats such as portable document files (PDF), audio and video files, presentation formats like PowerPoint files.

Network connection. Web-based applications run smoothly on native connections for intranet and Internet purposes. The data transmission is being sent efficiently to the common network infrastructure. However, due to file size, each transmission and rendering vary. Intranet connection transfers files from one machine to another efficiently; however, Internet connection runs on a different network configurations and connections where the data arrival and web-layout rendering differ according to ISP bandwidth allocations to every connected machine or client. Despite the geographical locations, other employees who are assigned in other campuses will be able to retrieve files using the Virtual Private

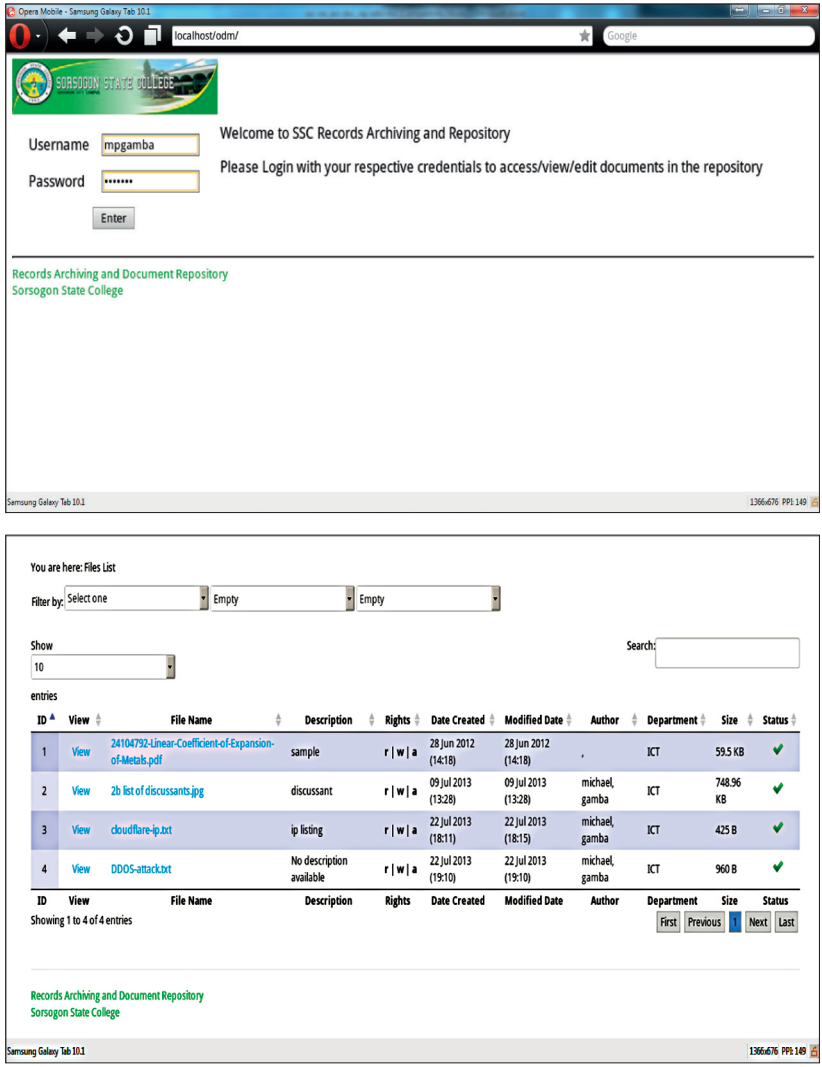

Figure 4. Screenshots for Mobile/Tablets Integrations

Network (VPN) to have a secure connection where the archiving interface will be visible only to the employees of the institution. The Hamachi and TeamViewer are the two open/free VPN software that have been tested for the web app. The latter works smoothly on its VPN connection due to pure IPv4 addressing, while the other software works on both IPv6 and IPv4 creating a 6to4 tunnel; thus, data transmission is being related in this method from one to the other end of the network which results to slow rendering of web layout of the web-based application interface (See Figure 5).

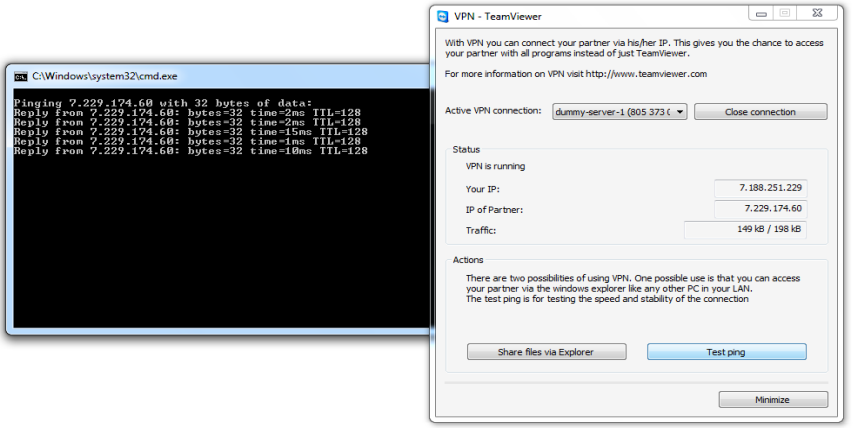

Figure 5. TeamViewer VPN Screenshots
for Web App VPN Connection

Repository. Digital documents repository will be stored on the same machine where the web app is located during its development period. This way of storing data creates a faster retrieval of the actual file from the disk location to the presentation layer of the interface, but for added security measures, these digital documents will be stored on a separate hardware for the actual deployment on the production server. The file server only authenticates the web server machine during its file retrieval. Other connections that are outside the network segment or unauthorized device will not be allowed for any form of file access. The diagram below (Figure 6) shows the simplified network connections and data flow.

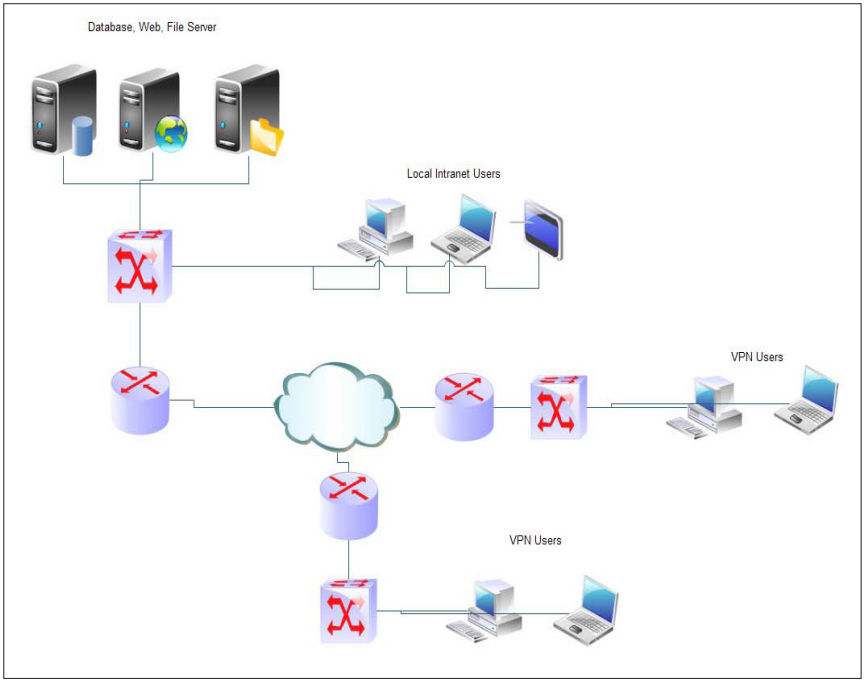

Figure 6. Simplified Repository Server and Network Connection

The greater flexibility of the networked data from one place to another makes it easier for every personnel to grab the updates at real time, together with the services being shared.

CONCLUSIONS

The study developed a records archiving and document repository applications which convert the common documents for Sorsogon State College' regular activities and transactions. The developed application overcomes the common issues on platform dependency and compatibility on operating system core files. The application had been built on top of these operating systems that were publicly available. It only requires a functional web-browser on the client side to render its interface, which is more cost effective and efficient. Hence, most of the desktop, laptop, and mobile devices contain web browser applications it would be easier to deploy the application since its design and development were built on

top of these web browsers. It also lessens workload on network configurations for its intranet deployment. Since, users can access the general user interface of the application, a copy of it will be downloaded and rendered to their web servers, thus, data contents that returns to the file server were only the requests, but not the whole content of the application itself.

The application can immediately address the increasing demand and simultaneous request of data to and from the other campuses. The output will be rendered or accomplished at their work stations.

Private servers require different methods of employing a direct access from its clients to preserve data integrity and isolation from public viewing. The developed applications used the public technology such as open/free virtual public network (VPN) to layer the connection to other stakeholders and to access directly the private servers which is geographically located. As a result, this developed system for records archiving and digital document repository will minimize the amount of time spent for SSC personnel's travel to and from their respective workstation and will maximize personnel's productivity for their immediate data updates.

TRANSLATIONAL RESEARCH

The limitations of the paper-based materials to be stored for future reference are being degraded for its usability due to several environmental factors. This study proposed to establish a repository for digital documents which can be used to propagate documents for various purposes in different offices of the agency. Due to the presence of natural calamities in the country's settings, this study was proposed as archiving software for those documents that were official or historical in nature in the agency. This system can be effectively managed and deployed under intranet or local area network since the output of this study is in the web-based format, which can run in all types of web-browsers, but effectively deployed on Mozilla Firefox and Google Chrome. This software has not been fully tested on wide-area network for offices that were separated in distant geographical locations. It does not limit on notebook and desktop computers, it also extends its output rendering on other platforms such as mobile devices that either run on android or other operating systems as long as there is a web-browser installed in it.

LITERATURE CITED

Askhoj, J., Nagamori, M., & Sugimoto, S. (2011, February). Archiving as a service: a model for the provision of shared archiving services using cloud computing. In *Proceedings of the 2011 iConference* (pp. 151-158). ACM.

Buchanan, S., & Richardson, K. (2012). Representation through Documentation: Acquiring Student and Campus Life Records through the Bruin Archives Project. *American Archivist*, *75*(1), 205-224.

El-Bakry, H. M., & Mohammed, A. A. (2009). Optimal Document Archiving and Fast Information Retrieval. *The International Journal of Computer science, and Engineering*, (2), 108-121.

Esser, D. (2013, January). Cooperative and Fast-Learning Information Extraction from Business Documents for Document Archiving. In *On the Move to Meaningful Internet Systems: OTM 2013 Workshops* (pp. 22-31). Springer Berlin Heidelberg.

Giesecke, J. (2011). Institutional repositories: Keys to success. *Journal of Library Administration*, *51*(5-6), 529-542.

Gomes, D., Miranda, J., & Costa, M. (2011). A survey on web archiving initiatives. In *Research and advanced technology for digital libraries* (pp. 408-420). Springer Berlin Heidelberg.

Kim, J. (2011). Motivations of faculty self-archiving in institutional repositories. *The Journal of Academic Librarianship*, *37*(3), 246-254.

Kohl, C. D., Bruns, I., Freudigmann, M., Scharf, G., Schmücker, P., Schwarz, G., & Semler, S. C. (2013). GCP-compliant digital archiving of paper-based patient records of clinical trial subjects: a key issues paper. *Clinical Investigation*, *3*(5), 451-465.

Krishnamurthy, M., & Kemparaju, T. D. (2011). Institutional repositories in Indian universities and research institutes: a study. *Program*, *45*(2), 185-198.

Nadeem, A., Yousaf, M. H., & Habib, H. A. (2010, August). Management information system for documents archiving and organization security. In*Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on* (Vol. 6, pp. V6-1). IEEE.

O'Reilly, T. (2005). What is web 2.0, design patterns and business models for the next generation of software. Retrieved December 1, 2008.

Pennock, M., & Kelly, B. (2006, May). Archiving web site resources: a records management view. In *Proceedings of the 15th international conference on World Wide Web* (pp. 987-988). ACM.

Seadle, M. (2012). Archiving in the networked world: authenticity and integrity. *Library Hi Tech*, *30*(3), 545-552.

Shiman, L. G., & Andrews, M. J. (2001) *U.S. Patent Application 09/874,704*.

Wu, H., & Gordon, M. D. (2004) Research and Development in Information Retrieval. *Proceedings of the 27th Annual International ACM SIGIR Conference*, 518-519.

Yoo, K. (2014). Intelligent and Pervasive Archiving Framework to Enhance the Usability of the Zero-Client-Based Cloud Storage System. *International Journal of Computer Science & Information Technology*, *6*(1).