Assalamualaikum,

Alhamdulillah, first of all we would like to thank Allah S.W.T as finally we are able to accomplished our task in publishing of TINTA about LEADERSHIP IN INFORMATION MANAGEMENT. We also would like to say thank you to our family and friends for supporting us throughout the time when we were doing this task.

Furthermore, we also would like to express our gratitude to our advisor Mr Azmi bin Abdul Rahman for all the advices, guidance that he had given throughout the task and we also would like to say thank you again for all the moral supports that he had given to us in completing this publishing of TINTA.

Now we understood how the publication process and this process is not simple like what we think.

Last but not least our big thanks to all the contributors that contribute their article for our TINTA publication. Without them there is no meaning of this TINTA publication.

Again, we would like to thank everyone who had involve in helping us finishing the assignment and all the supports that they had given to us and happy reading.

Editorial Team,

TINTA (Leadership in Information Management)
ISI KANDUNGAN

BAHAGIAN A

What is leadership ?..............................................................1
By Nur Iliana binti Baharudin Shah

Good leader and effective leadership..................................6
By Hafizah Binti Md Aris, Norsyuhada Binti Mohd Hanafi & Shantie Binti Mohamad

The Challenges and Importance of Good Role Leadership in Information Management......14
By Akilah Nadirah Binti Mohd Roszelin, Nur Anis Syuhada Binti Ab Malek & Umi Nazirah Binti Zainal

Effect of good leadership..................................................20
By Nurul Syahirah binti Md Salleh, Nur Syazwani binti Bakar, Norli Ezzati binti Mohd Abu Hassan AlShari

Leadership in public libraries.............................................24
By Aida binti Kamaludeen, Amir bin Muhamad Shafie, Muhamad Hafiz bin Abdullah.

Who's who in the library....................................................28
By Nurul Syahirah binti Md Salleh, Nur Syazwani binti Bakar, Norli Ezzati binti Mohd Abu Hassan AlShari

En. Shaharom TM Sulaiman..................................................34
By Akilah Nadirah binti Roszelin, Shantie binti Mohamed and Mohamad Hafiz bin Abdullah

An Untold Leadership Story...............................................39
By Kevin Anak Ong

Leading to the Innovation in a public library..........................45
By Nur Ain binti Abd Wahab

Signage and Wayfinding in Library Planning and Design...............53
By Ahmad Zulhilmi bin Azmi & Rosidah Ismail

Preservation and Conservation...........................................59
By Afiq bin Mohd Shahari

Web Archiving and Archive Management................................66
By Syuhaidah Abdul Rahman

Laporan Program Annual Grand Meeting Gen-LIS 2014................77
By Hafizah bt Md Aris & Nur Anis Syuhada binti Ab Malek

Laporan Lawatan Percetakan Nasional Malaysia Berhad (PNMB)..........83
By Umi Nazirah bt Zainal
Book Review .................................................................................................................. 91
(BIBLIOHOLISME: Menelusuri Pesona Dunia Buku dan Pencintanya)
By Mohamad Hafiz bin Abdullah

Contributors .................................................................................................................. 94
Web Archiving And Archive Management

By
Syahidah Abdul Rahman
Faculty of Information Management
UiTM

ABSTRACT

This term paper is discussed about the web archiving and archive management. Web archiving has been gaining interest and recognized importance for modern societies around the world. However, for web archivists it is frequently difficult to demonstrate this fact, for instance, to funders. This study provides an updated and global overview of web archiving. Considering the complexity and large amounts of data involved in web archiving, the results showed that the assigned resources are scarce. Besides that, this paper also discuss about the current issues and challenges that are face in web archiving and also include the best recommendation and solution about that matter. While for archive management, it is focused on the roles of archivist and the importance of archive institution as a permanent storage for records and documents to the community.

Keywords: Web archiving, Archive management, Digital storage

INTRODUCTION

According to International Records Management Trust (2012), archives are records, usually but not necessarily non-current records, of enduring value selected for permanent preservation and will normally be preserved in an archival repository. Archives are those records that are worthy of permanent retention because of their enduring value as evidence or for research. Archives are an elite body of records. They provide a reliable and authentic knowledge base, enabling the past to be reconstructed and understood. Without archives, the past would remain largely unknown. By documenting the significant decisions, transactions and events of political, social and economic life, archives serve as the essential link in the chain of human history. Archives are preserved in and managed by specialist archival institutions where they are safeguarded and made available for use. The word “archiving” often refers to the process of storing physical objects, generally though not exclusively paper-based, in a physical location, such as a room or a building, to maintain that object’s physical integrity and its intellectual context as could be represented by other objects within the archive (Seagle, 2006). The term “archiving” also can be substituted for preservation provided this definition remains, but “archiving” is usually interpreted within the computing industry simply to indicate that something has been stored and is no longer immediately accessible.
The uses of archives

Few records are recognised as having archival value at the point of their creation. The perceived value and use of the bulk of records change over time (IRMT, 2012). A file created this week may have great archival value in twenty-five years, but it can be difficult to discern that value at this early stage in the record's life. Furthermore, the value of records differs between the different communities of users. For the creator or original user, the usefulness of the records is likely to diminish once the record has fulfilled its initial purpose. Managers and auditors, for example, use records to access performance and accountability. To them, the administrative value of records ends when regulatory requirements have been met. The majority of records will be destroyed once their business function has ended and there is no further need to retain the evidence they contain (IRMT, 2012). Citizens also revere archives associated with great events or persons in history. Some documents will have an enduring value to the citizen because they enshrine evidence of his or her legal rights or family history. Record keepers and archivists must be able to prove that the essential characteristics. In spite of the different uses to which archives are put, all the different perceived values of archival materials rest upon a common foundation. In order for a record to provide valuable research information, its value as evidence and its integrity, authenticity, and meaning must be retained intact. Organisations and individuals can be exposed to business and legal liability if they neglect essential record-keeping requirements (IRMT, 2012). However, the long-term consequences of neglect can be even more serious. Unique and irreplaceable evidence of the past may be lost forever if bodies of records are destroyed without being appraised for their enduring value or when records are damaged through neglect or mishandling. Record keepers and archivists must be able to prove that the essential characteristics of records as evidence have been protected and preserved over time. The value of archives as authoritative evidence depends upon the quality of their custody and care from the time they were first created and used. The archivist is perfectly placed to understand this requirement. At any point in their life, whether in the custody of their creators or an archival institution, records are unusable or valueless if their context and authenticity has not been safeguarded (IRMT, 2012). Without context and authenticity, the evidence contained in archives cannot be relied upon or fully understood. Restoring the order and context of archives is also an important part of the work of archivists. Inevitably, accidents and mishaps occur, responsibilities change and organisations evolve. Sometimes records are partially or completely destroyed without regard to their enduring value. At other times, the original order of records is disturbed, making it difficult if not impossible to place the records into context. Private papers or the records of small or defunct organisations may also come to the archivist in a disorderly and incomplete state (IRMT, 2012). Hence archivists and records professionals must sometimes reconstruct the past from surviving records and other materials.
Web archiving

The history of the Web is still in its infancy. Nevertheless, the Web has already had a great impact on our information space. Nowadays, many websites are culturally valuable and many important academic works have been disseminated on the Web although it is obvious that the number of inactive URLs is gradually increasing (Lawrence et al., 2001; Dellavalle et al., 2003; Koehler, 2004). There is a risk that future generations might not be able to access the cultural heritage accumulated by their predecessors (Lyman, 2002). Web archiving is a way to counteract these fragile characteristics of the Web and to ensure long-term access to information on the Web. At present, many institutions, especially national libraries, are attempting to construct archiving infrastructures with the help of domestic and international cooperation. In addition, as Joint (2006) suggest, a website collection, unlike one of e-monographs or e-journals, would represent a unique cultural heritage for each country. Assessing the size of the Web is a difficult task, and many attempts to provide a reliable estimate of it have been made so far with limited success. This paper will review on outline major changes the Web has introduced and discuss their impact for web archiving.

CURRENT ISSUES/ CHALLENGES

Authorship Revolution

The email and blog phenomenon is the most recent illustration of this revolution which the first Web browser designed and coded by Tim Berners Lee included an authoring tool, which he considered to be an essential piece of the new system. Blogs have been used for the distribution of ideas such as publishing diaries, personal thoughts, webpage scraps and etc. Blogging is a new form of diary writing, and expands on the concept of the diary. The value of blogs should not be greatly different from that of diaries, the historical antecedents of blogs (Gillies & Gaillau, 2000; Berners-Lee & Fischetti, 2000). Despite the subsequent omission of authoring tools from Web browsers, the Web has continued to offer an open publishing platform with global accessibility and continuous updating capacity. This has dramatically changed the setting for publication, allowing almost anyone to bypass the traditional publishing actors and reach direct access to a potentially unlimited audience. The eventual impact of this change remains to be seen, but several consequences for archiving are already tangible. The first important change is the end of an object’s stability, with obvious impacts for archiving—an activity that in essence consists of capturing the state of an object at a point in time. The Web offers the ability to update content at any moment without notification, if additional notification mechanisms like Really Simple Syndication (RSS) protocol feeds are not in place, which poses a great challenge for archivists. Revisiting pages consumes resources, even if heuristics can be found to alleviate this process. Choice of an appropriate frequency for capture can be problematic because, to be efficient, it should be done at the page level in most cases. It is indeed equivalent to assessing the probability of losing some intermediary updates between two captures (Donghee, S., Sue, Y.S., & Sugg, M.K, 2011).
Content Shaping

In addition to the change in the publication process, an important shift has occurred in the nature of documents themselves. The proliferation of citations that the hypertext environment allows induces a tremendous tendency toward dispersal of content, which archivists have to take into account in their approach. Web documents at the page level (but also the site level) hardly ever make sense alone. They are mingled in a larger document network that forms what Nelson named a “docuverse” (Nelson, 1992). From this perspective, archiving means extracting slices of the Web that constitute a whole meta document according to Landow (1997) that is, spatially sampling the Web and making decisions each time regarding the exact perimeter of what to include, being aware that with time non-inclusion means loss. For example, does archiving a site mean leaving out any document linked outside of its domain? If not, to what depth should external links be followed? There is no general answer to these questions, only specific ones based on the goal driving the archiving. Choices also have to be made concerning what characteristics or functionalities are to be preserved. When a site is not primarily a collection of static pages, an archivist may focus on the interaction of functionalities which is not only for navigation and more generally the experience the site provides in the archival context.

Convergence

It is worth noting that the Web is not only a platform absorbing previously existing Internet applications such as mail, FTP, news and as well as non-Internet based applications like database, document repository, and various information systems, but it also tends to be an entry point for almost everything today. This is a clear consequence of the design adopted for Uniform Resource Identifiers (URI), which Tim Berners-Lee insists is the most important standard of the Web (Berners-Lee & Fischetti, 2000). The prefix, which means the use of the Domain Name Server (DNS) system for host naming, and the flexibility offered for Webmasters regarding the right part of the URI, together make URI a powerful unifying standard. But for archivists this means almost everything can end up in their nets. If they want to focus on published material in the traditional sense of the word, they might want to filter online forums, for instance, or avoid diving into huge databases. Clues can be used for limiting the archiving, using, for instance, URI pattern detection (this has long been the case with search engines avoiding any dynamically generated content based on URI-embedded queries). This can extend to filtering content on the fly or during post-processing.

Technique

Even when the target is clearly identified and delimited, content acquisition can be an issue. Automatic tools for content gathering such as crawlers which also called spiders allow massive content acquisition at relatively low cost. With standard desktop computers and a Digital Subscriber Line (DSL) connection, it is possible today to retrieve millions of documents per week, even per day. Crawlers are also powerful and systematic tools for exploring the Web and discovering new sites through links even when starting from a very small set of seed sites. There are severe crawler limitations, however, when it comes to finding a path to certain types of documents.
First, access to sites or parts of sites can be restricted with password or Internet Protocol (IP) authentication. In this case, getting authorization is needed. Second, the coding technique used to implement links can be hard to interpret for crawlers. This can be the case when scripts use contextual elements or when the code is opaque which are executable, server-side code and etc. Crawlers are getting better at link extraction but still face some limits. Finally, a nontrivial interaction from the user can be required and that is, more than a click. This is usually the case when entering a query is required to access some portion of content. Content acquisition in this situation entails a case-by-case assessment, and adapted actions must be taken. This can be limited to entering new parameters for the crawler or downloading directly page by page some part of the site. In many cases still, nothing can be done remotely, and getting the content through the hypertext transfer protocol (http) interface is not possible. In these cases, pursuing direct contact with the producer is unavoidable, which is extremely time consuming compared to direct online capture. To summarize this quick overview of the situation, this can be observe that the extraordinary extension of opportunity the Web offers for producers results in a corresponding increase in difficulty for archivists. Therefore, there must have solution about this matter (Masanes, 2005).

RECOMMENDED SOLUTIONS

The solutions for the web archiving in email and blogs are creating a retention policy for e-mail is often challenging in the organizational environment, and is the major reason behind delays on taking action for e-mail preservation. Some scholars have recommended archiving all e-mail messages (Dando-Collins, 2006; Perry, 2008), while others have attempted to develop preservation tools with better functionalities. Even though many institutions have encountered difficulties in interpreting and implementing regulations for retention and appraisal processes, these legal requirements still provide a mechanism to preserve e-mail in organizations. Yet the preservation of personal e-mail relies on the awareness of their creators and the decisions they make. While the method to save and preserve the blog, mainly in PDF or image formats, bloggers must search for tools on their own such as use freeware developed by third parties because no other options are apparent to them.

Quality

The quality of a web archive can be defined by the completeness of material archived within a designated perimeter and being able to render the original form of the site, particularly regarding navigation and interaction with the user. Graphically, completeness can be measured horizontally by the number of relevant entry points found within the designated perimeter and vertically by the number of relevant linked nodes found from this entry point. Usually, entry points are site home pages, and links can direct the user either to a new entry point which means another site or to elements of the same site. This is the case for site-oriented archiving. In some cases, however, verticality is limited to in-line documents for example, images associated with a page, and the collection is just organized horizontally ignoring the site level. This is the case, for instance, for pure topic crawling where nodes are not included based on their
The web archiving can be preserved through Pandora. Archive-It is a hosted web archive service provided by the Internet Archive. The first Australian organisation to use this service has been the Asian Collections section of the National Library. It is used to gather collections of overseas web sites recording particular social and political events, as it is not expected that any other regional organisation will fulfill this role. The hosted option was chosen as it was perceived that it could be a quick and easy way of gathering and storing collections, and not requiring technical skills or big tranches of staff time. This has proved only in part to be true as it was quickly noticed that to successfully create collections takes far more curatorial time than was initially envisioned. The selection of websites to crawl is an often misunderstood activity and can take up surprisingly large amounts of time. While using Archive-IT has benefits in that archivist do not need to worry about hosting and preserving gathered content, there are also some major drawbacks about not being in control of the archival and display functions. Archive-IT allows for a set of the chosen seed URLs to be gathered, however once the gatherer has gone out and gathered, or not gathered, the files, there is no way to manually fix up any broken or missing content, as can be done with their own system. Similarly there is no real control or ownership of the display process, so that for instance a link to a seed URL which has failed to gather will still appear within a collection. Another drawback is that if they can discontinue the annual subscription to the service on collections are dispersed back into the general internet archive pool of content. Notwithstanding these issues the library plans to continue to archive using this method (Crook, 2009).

Lots of Copies Keep Stuff Safe (LOCKSS)

LOCKSS is open source, peer-to-peer software that functions as a persistent access preservation system. According to Seadle (2006), LOCKSS archives the whole Web site that it crawls as a bit stream. This is the digital equivalent of the archival principle of preserving the original object in its context without rebinding, retouching, or other transformations that later generations may regret. Information is delivered via the web, and stored using a sophisticated but easy to use caching system. LOCKSS software allows the libraries to collect, store, preserve, and archive authorized content locally. The local copies serve as back-ups and can be accessed when the publishers’ site becomes unavailable. The LOCKSS system uses a crawler to collect e-journal content from the publishers’ web sites as it is published. With written and machine-readable permissions from the publishers are required for this. Publishers are encouraged to grant libraries legal permission to cache and archive their content via language in licenses or terms and conditions. Helen Hockx-Yu emphasizes that a number of publishers have participated in LOCKSS testing, for example Blackwell, Project Muse, British Medical Journals Publishing, Oxford University Press. The additional Mellon funding is intended to support the next stage of LOCKSS, to manage content as bibliographic entities rather than as web-addressed files (Golnessa, 2008).
Conclusion

In a nutshell, through this paper, it can show the challenges associated with web archiving require consideration and appraisal of a variety of approaches that can complement each other and allow better global efficiency for preservation of web content. The challenges are first is about authorship revolution which is about the email and blog phenomenon where people are usually use this medium to save their records and documents without realizing the importance of preservation their email and blog. Second is about content shaping where the proliferation of citations that the hypertext environment allows induces a tremendous tendency toward dispersal of content, which archivists have to take into account in their approach. The third is about convergence which is not only a platform absorbing previously existing Internet applications such as mail, FTP, news and as well as non- Internet based applications like database, document repository, and various information systems, but it also tends to be an entry point for almost everything today. This is a clear consequence of the design adopted for Uniform Resource Identifiers (URI). Next is technique that is focus on crawler. Crawlers are also powerful and systematic tools for exploring the web and discovering new sites through links even when starting from a very small set of seed sites. There are severe crawler limitations, however, when it comes to finding a path to certain types of documents.

While the recommendation and solutions that are suggested are for email and blogs it using preservation tools with better functionalities. Even though many institutions have encountered creating a retention policy for e-mail is often challenging in the organisational environment, and is the major reason behind delays on taking action for e-mail preservation through archiving all e-mail messages, while others have attempted to develop difficulties in interpreting and implementing regulations for retention and appraisal processes. Apart from that the recommendation is about quality. The quality of a web archive can be defined by the completeness of material archived within a designated perimeter and being able to render the original form of the site, particularly regarding navigation and interaction with the user. Usually, completeness can be measured horizontally by the number of relevant entry points found within the designated perimeter and vertically by the number of relevant linked nodes found from this entry point. The other recommendation is web archiving can be preserve through Pandora. Archive-It is a hosted web archive service provided by the Internet Archive. It is used to gather collections of overseas web sites recording particular social and political events, as it is not expected that any other regional organisation will fulfill this role. The hosted option was chosen as it was perceived that it could be a quick and easy way, of gathering and storing collections, and not requiring technical skills or big tranches of staff time. Besides that, Lots of Copies Keep Stuff Safe (LOCKSS) is also one of the recommendations. LOCKSS is open source, peer-to-peer software that functions as a persistent access preservation system where the whole Web site that it crawls as a bit stream. This is the digital equivalent of the archival principle of preserving, the original object in its context without rebinding, retouching, or other transformations that later generations may regret.
Last but not least, through this term paper, its hope that this work will contribute to laying the ground for Web archiving appraisal and adaptation in the future and the archive institution and the community must work together to make their historical and important records are preserve and for the sake of next generation to know where they are come from and be able to continue to appreciate and know the importance of archival materials and keep them securely whether in a soft copy such as web archiving or in a hard copy for further use.

References


Dando-Collins, S. (2006). The e-mail, the whole e-mail, and nothing but the e-mail. Records Management Society Bulletin, 132, 5.


### Contributors

<table>
<thead>
<tr>
<th>Name</th>
<th>Title and Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AZMI BIN AB RAHMAN</strong></td>
<td>Lecturer of Faculty of Information Management, Universiti Teknologi Mara Kampus Puncak Perdana</td>
</tr>
<tr>
<td><strong>SHAHAROM TM SULAIMAN</strong></td>
<td>Lecturer of Faculty of Information Management, Universiti Teknologi Mara Kampus Puncak Perdana</td>
</tr>
<tr>
<td><strong>KEVIN ANAK ONG</strong></td>
<td>Business Development Executive, Regalia Records Management</td>
</tr>
<tr>
<td><strong>NUR AIN BINTI ABD WAHAB</strong></td>
<td>Bachelor of Science (Honors) Records Management, Universiti Teknologi Mara Kampus Puncak Perdana</td>
</tr>
<tr>
<td><strong>NUR ILIANA BINTI BAHARUDIN SHAH</strong></td>
<td>Work as PEGAWAI ARKIB SEKSYEN PERANCANGAN STRATEGIK &amp; ANTARABANGSA DI ARKIB NEGARA MALAYSIA</td>
</tr>
<tr>
<td>Name</td>
<td>Degree</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>NORSYUHADA BINTI MOHD HANAFI</td>
<td>BACHELOR OF SCIENCE (HONORS)</td>
</tr>
<tr>
<td>AIDA BINTI KAMALUDEEN</td>
<td>BACHELOR OF SCIENCE (HONORS)</td>
</tr>
<tr>
<td>AMIR BIN MUHAMAD SHAFIE</td>
<td>BACHELOR OF SCIENCE (HONORS)</td>
</tr>
<tr>
<td>NORLI EZZATI BT MOHD ABU HASSAN ALSHARI</td>
<td>BACHELOR OF SCIENCE (HONORS)</td>
</tr>
<tr>
<td>NUR SYAZWANI BINTI BAKAR</td>
<td>BACHELOR OF SCIENCE (HONORS)</td>
</tr>
<tr>
<td>UMI NAZIRAH BINTI ZAINAL</td>
<td>BACHELOR OF SCIENCE (HONORS)</td>
</tr>
</tbody>
</table>
AKILAH NADIRAH BINTI MOHD ROSZELIN
BACHELOR OF SCIENCE (HONORS)
LIBRARY AND INFORMATION MANAGEMENT

SHANTIE BINTI MOHAMAD
BACHELOR OF SCIENCE (HONORS)
LIBRARY AND INFORMATION MANAGEMENT

HAFIZAH BINTI MD ARIS
BACHELOR OF SCIENCE (HONORS)
LIBRARY AND INFORMATION MANAGEMENT

NUR ANIS SYUHADA BINTI AB MALEK
BACHELOR OF SCIENCE (HONORS)
LIBRARY AND INFORMATION MANAGEMENT

NURUL SYAHIRAH BINTI MD SALLEH
BACHELOR OF SCIENCE (HONORS)
LIBRARY AND INFORMATION MANAGEMENT
UNIVERSITI TEKNOLOGI MARA KAMPUS PUNCAK PERDANA

MOHAMAD HAFIZ BIN ABDU'L-LAH
BACHELOR OF SCIENCE (HONORS)
LIBRARY AND INFORMATION MANAGEMENT
ROSIDAH BT ISMAIL

KETUA BAHAGIAN PUSAT REKOD
DEWAN BANDARAYA KUALA LUMPUR

AHMAD ZULHILMI BIN AZMI

PEMBANTU PERPUSTAKAAN
PERPUSTAKAAN SULTANAH ZAHARIAH
UTM KUALA LUMPUR

SYAHIDAH ABDUL RAHMAN

BACHELOR IN LIBRARY SCIENCE
UNIVERSITI TEKNOLOGI MARA

AFIQ BIN MOHD SHAHARI

BACHELOR IN LIBRARY SCIENCE
UNIVERSITI TEKNOLOGI MARA