May 2019

LOCKSS, CLOCKSS & PORTICO: A LOOK INTO DIGITAL PRESERVATION POLICIES

Ubaid Ullah Shah
shahubaid7@gmail.com

Sumeer Gul Dr.
Department of Library and Information Science, University of Kashmir, India

LOCKSS, CLOCKSS & PORTICO: A LOOK INTO DIGITAL PRESERVATION POLICIES

Abstract

Digital preservation is the process of preserving information in the digital form. By digital form we mean a non- print form. Information is moving from print to web (non-print) and it has become easy for masses especially the scholarly population to access information on the web from anywhere 24*7. Since digital information is more prone to damage and loss than print information, therefore, it is imperative to protect access to this information not only for the benefit of present but for future generation as well. Various initiatives and policies are currently in place to ensure the long term preservation of digital information. LOCKSS, CLOCKSS, and PORTICO are the three major digital preservation initiatives aimed at protecting and preserving the digital content for its long term access and use. The current study provides an overview of how these three initiatives started and how these policies function in collaboration with publishers and libraries for long term digital content preservation.

Keywords: *LOCKSS; CLOCKSS; PORTICO; Digital Preservation*

INTRODUCTION

Earlier information used to be available in print form and libraries used to purchase it from the publishers and had a lifelong access to the material but now information is migrating to web, forcing the libraries to rent the information from publishers instead of purchasing it without any guarantee of forever or long term access **(Maniatis, Roussopoulos, Giuli, Rosenthal & Baker, 2005).** Scholars have also become more dependent upon the electronic versions of scholarly journals, because of their ease of availability and access but the long-term preservation of these resources has become urgent need of the time as the information in digital form is more prone to damage than in print form because of many reasons like technological obsolescence, media failure, hardware and software failure, network failure etc **(Fenton, 2006)**. Therefore digital preservation to libraries is indispensable. Digital Preservation is simply the maintenance and management of digital content or objects. It refers to the preservation of material which was originally created in digital form(Born Digital) as well as the material which existed in analog form but was later converted into digital form for preservation and access purposes **(Arora, 2009)**. Digital preservation can be defined as the "cumulative actions undertaken by an organization or individual to ensure that digital content is usable across generations of

information technology" **(McGovern, 2009).** For digital preservation of material, certain policies and programs are required. A digital preservation policy is the authorization for an archive to preserve the digital records through a controlled (structured) and managed digital preservation strategy. There are many software systems, policies, and archives developed around the globe for the purpose of preserving the digital content; the three main important initiatives are LOCKSS, CLOCKSS, and PORTICO.

LOCKSS: LOCKSS, an acronym for Lots of Copies Keep Stuff Safe, is the free and open software developed under the auspices of Stanford University about 20 years ago. It works on the principles like replication, format migration and repair through a polling mechanism. When LOCKSS is downloaded and installed on an individual system, it automatically becomes part of a LOCKSS network comprising of hundreds and thousands of nodes. Each node connected to network share information with one another as a result of which multiple copies of information gets produced at different geographical locations. LOCKSS includes a unique polling and repair mechanism. If particular information on a node gets deleted or reports a difference in comparison to other nodes, it is automatically polled out odd in the network and the deleted information is replaced or the difference found is neutralized thereby keeping the data secure for long. LOCKSS also possesses the feature of format migration i.e. if the information is available in the format which is no longer valid; LOCKSS automatically changes it to the usable or current format. "Through LOCKSS, requests for material are sent to the publisher, and if the content is not retrieved for any reason, the LOCKSS copy is provided with one exception of dynamic content i.e advertisements or graphics that change with each screen display, remains static". There are near about fifty publishers who are currently participating in LOCKSS program, and most of these have agreed to allow their subscribers to use LOCKSS as a backup system **(Rutgers University Libraries, 2007)**.LOCKSS is a "light Archive" i.e. it provides access to archived information even in case of temporary loss of access to publisher or publishers website **(Swoger, 2014)**. Though LOCKSS is open and free software, its members have to join an alliance called LOCKSS alliance. The annual fee for LOCKSS ranges between INR 77000-762000 **(Rutgers University Libraries, 2007)**.

CLOCKSS: CLOCKSS stand for Controlled Lots of Copies Keep Stuff Safe is an extended version of LOCKSS. Presently it has 12 active nodes and covers over 30 million journal articles and 75,000 books. It is an archiving service which ensures that the content will be made

available to users in spite of economic, technological, political and environmental disturbances. It works on the same principle of replication and polling odd one out mechanism as that in LOCKSS but unlike LOCKSS (a light archive) it is a dark archive, which ensures access to information in trigger situations only. According to CLOCKSS archive, trigger event is when content is unavailable from any publisher for at least six consecutive months and there is no information about transfers of rights to the board at the time of the determination **(Dyck, 2019)**. Trigger event mainly occurs because of the following reasons:

- Publisher no longer exists
- A publisher has stopped publishing particular information
- A publisher has stopped giving access to some or all back issues
- Catastrophic events: Loss of information due to the occurrence of events like floods, earthquake etc

In its 12 years, CLOCKSS has triggered 53 journals comprising 13,000 articles, which is a very small percentage of the 2000+ journals and 30 million articles in the Archive. CLOCKSS provides Open Access to triggered content **(CLOCKSS, 2019)**.

PORTICO: Portico is a digital preservation service for electronic journals, books, and other content **(Morrissey et al., 2010).** The project was initiated by JSTOR with a grant from The Andrew W. Mellon Foundation. Initially, the project was focused on designing and prototyping content handling and archival systems, crafting potential archive service models, testing possible models with libraries and publishers, and a drafting business model able to support a long-term archival effort**.** In 2004, the project was undertaken by Ithaka- the U.S based not for profit organization, whose main aim is to help the academic community. Ithaka brought librarians and publishers together and discussed various issues and challenges which come across while developing an archive. Finally after a lot of discussions and cooperation Portico was shaped and launched in 2005 **(Fenton, 2006)**. Portico on receiving the source file from the publisher passes it through its content-normalization process. It receives the source files in the form of publishers' proprietary XML and SGML format and it normalizes them to format which Portico and other digital preservators have adopted, then both the original as well as the normalized content is preserved in the archive **(Morrissey et al., 2010)**. PORTICO just like CLOCKSS is a "dark archive" i.e. it provides access to archived content in case of any trigger event, which includes the following:

- Publisher no longer exists.

- Publisher ceases to publish a title

- Publisher no longer offers access to back issues

- Failure of a publisher's delivery platform because of any catastrophic event

- Currently, PORTICO serves as an archive for 11,954 electronic journal titles, 65,986 e-book titles, and 39 digitized historical collections

Conclusion

Preserving digitized content is more complex than preserving a document in print form. Since the digital world is a place where technology is rapidly changing and innovations are taking place frequently, there are always chances of information being lost or being no more accessible. Under such an environment, we are in need of strategies, programs, and policies leading to conservation and protection of the digitized content. Digital preservation policies and initiatives like LOCKSS, CLOCKSS and PORTICO and many more have reduced a lot of stress from the librarians in making the content or information safe and accessible to not only present but also to future generations in spite of ever-changing technology. Our societies are in need of more preservation policies as the qualitative scholarly information is generated on a large pace around the globe and the presence of few policies is not sufficient. Thus, it is incumbent over libraries and publishers that they should come together in mutual interests of the society and work hand in hand to ensure preservation and lifelong access to ever valuable and vulnerable digital information.

References

Arora, J. (2009). Digital Preservation: an overview. *Proceedings of Open Access to Textual and Multimedia Content: Bridging the Digital Divide, New Delhi, January 29-30, 2009*. New Delhi: INFLIBNET.

CLOCKSS. (2019). *CLOCKSS: The CLOCKSS Archive: A Primer*. Retrieved February 17, 2019, from https://clockss.org/wp-content/uploads/2019/02/CLOCKSS-Primer.pdf

Dyck, C. (2019). *The CLOCKSS Archive: Long-Term Preservation for Scholarly Content.* U.K: Ingenta. Retrieved from: https://www.ingenta.com/ingenta-connect-publisher-newsletter-london-book-fair%20edition/clockss-archive-long-term-preservation-scholarly-content/

Fenton, E. (2006). An Overview of Portico: An Electronic Archiving Service. *Serials Review*, *32*(2), 81-86. doi: 10.1080/00987913.2006.10765036

Maniatis, P., Roussopoulos, M., Giuli, T. J., Rosenthal, D. S., & Baker, M. (2005). The LOCKSS peer-to-peer digital preservation system. *ACM Transactions on Computer Systems, 23* (1), 2-50. doi:10.1145/1047915.1047917

Morrissey, S. M., Meyer, J., Bhattarai, S., Kurdikar, S., Ling, J., Stoeffler, M., & Thanneeru, U. (2010). Portico: A Case Study in the Use of the Journal Archiving and Interchange Tag Set for the Long-Term Preservation of Scholarly Journals. In *Proceedings of the Journal Article Tag Suite Conference*. Bethesda, MD: National Center for Biotechnology Information (US). Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK47087/

McGovern, N. Y. (2009). *Technology responsiveness for digital preservation: a model* (Doctoral dissertation. Gower St, Bloomsbury: UCL (University College London).

Rutgers University Libraries. (2007). LOCKSS, CLOCKSS, and Portico Potential Digital Archive Solutions for Rutgers. *Rutgers University Libraries Staff Resources*. Retrieved from https://www.libraries.rutgers.edu/rul/staff/collection_dev/reports/lockss-clockss-portico.shtml

Swoger, B. (2014). Preserving scholarly information: LOCKSS, CLOCKKS, and portico. *Scientific American*. Retrieved from: https://blogs.scientificamerican.com/information-culture/preserving-scholarly-information-lockss-clockks-and-portico/