

# Status of Dark Archives: A Study of Digital Preservation Initiatives

Rahat Khan\*, Atasi Sinhababu\*\*, Rupak Chakravarty\*\*\*, Nidhi Gupta\*\*\*\*

## Abstract

The purpose of this paper is to analyse the current status of digital preservation policy for libraries and publishers using dark archives. Another aim is to examine the status of open-access journals that are being indexed by DOAJ with regard to the adoption of LOCKSS, CLOCKSS and Portico during the process of digital preservation of scholarly research. The study also aims to examine the dark archiving policies of Open Access (OA) journals indexed by DOAJ, to know the status of the adoption of dark archive platforms, to know the preservation status and trigger events pertaining to dark archives and to examine the status of supporting libraries and participating publishers in the context of dark archives. The data for the study was extracted from DOAJ and the dark archive service providers. The data cleaning, filtering and consolidation were done for further analysis and interpretation of the data. Only 35% of open-access journals use preservation policy through dark archives, while 65% of OA journals do not use any dark archive policy. Findings show that 35,000 journal titles are preserved by CLOCKSS and 36,936 journal titles are preserved by Portico. On the other hand, in CLOCKSS, 13,000 articles have been triggered and no books have been triggered. In the case of Portico, 207 triggered events have occurred. The findings of the study throw light on the usability and applicability of long-term digital preservation (LTDP) of valued research with empirical evidences. Dark Archives facilitate trigger events shift the otherwise inaccessible (dark) into publicly accessible (light) so that the society at large can derive value and innovate.

**Keywords:** Dark Archive, Long-Term Digital Preservation, LOCKSS, CLOCKSS, Portico, DOAJ

## Introduction

Advancement in technology in the shape of tools and resources has made it possible to obtain the required information quickly. Libraries are subscribing to online resources to serve their patrons round the clock, augmented with remote access. Librarians, the publishing industry and non-profit organisations have a shared mission to maintain sustainable and affordable access to electronic scholarship for future generations. It is their responsibility to preserve original publications without altering them and to prevent loss caused by accidental corruption of content or unreadable formats when software or resources become obsolete. This can be achieved through digital preservation (DP), involving the practice of ensuring that digital information remains accessible and usable over time. DP includes ensuring that digital files are kept safe from corruption and degradation and that any necessary migration to new file formats and media is carried out in a way that preserves the original information (Malak, 2022). It is a process of preserving the integrity of electronic documents over time, ensuring that their legal and evidentiary value remains unaltered for as long as necessary. Digital preservation consists of the processes aimed at ensuring the continued accessibility of digital materials. To do this involves finding ways to re-present what

\* Librarian, LRC, Female Library, Prince Mohammad Bin Fahd University, Alkhoba, Kingdom of Saudi Arabia.

Email: rahatmlis2008@gmail.com; ORCID: <https://orcid.org/0000-0002-9575-0095>

\*\* Librarian, Amity University Punjab, Mohali, Punjab, India. Email: asinhababu@pb.amity.edu;

ORCID: <https://orcid.org/0000-0001-7036-0827>

\*\*\* Professor, DLIS, Panjab University, Chandigarh, India. Email: rupak@pu.ac.in;

ORCID: <https://orcid.org/0000-0001-5046-1663>

\*\*\*\* Librarian, SCD Government College, Ludhiana, Punjab, India.

was originally presented to users by a combination of software and hardware tools acting on data (Concept of Digital Preservation, 2019). For digital preservation publishers adopt different applications/technologies like DANS, e-DEPOT, LOCKSS, CLOCKSS, PORTICO and many more. The adoption of dark archives is a really expedient initiative to preserve the digital resources for libraries, publishers and open access (OA) journals. The contribution of LOCKSS, CLOCKSS and Portico in digital preservation reduces the risk of losing the digital resources for libraries, publishers and open access (OA) journals. These technologies provide a better approach to the preservation of digital resources in a systematic way. There are three approaches for archiving research, viz. Dark, Light and Dim. A dark archive is private, digital storage that is not accessible to the public (online readers). These serve as fail-safe repositories for storing and preserving published journal articles. Popular and commonly used dark archives include CLOCKSS, LOCKSS, Portico, etc. A light archive is publicly and openly accessible to all digital readers. Public archives provide you with various options for storing your scholarly articles, including institutional or university-run libraries, preprint servers, open archive databases and more. These repositories could be subject-specific, such as Cog prints (psychology, neuroscience, etc.), Engrxiv (engineering) and RePEc (economics). It also includes repositories built by universities and research institutions to host their research corpus and other forms of scholarly output. Examples: IIT Madras Repository, eScholarship by the University of California, etc. In Dim Archive, the model combines the previous two models. Some scholarly content is restricted to certain organisational custodians, but the rest is available to others, typically to a research community with a large readership. Dim Archives cannot be accessed by all internet users but can be released publicly if required or requested. The archival method is generally used when a journal needs to be preserved with limited access. This paper is an attempt to throw light on the policies and technologies for the long-term digital preservation (LTDP) of digital resources.

## Dark Archives: Standing the Test of Time

There are different options for archiving digital resources. The dark archives are an innovative approach toward

securing long-term access to digital scholarship in the event of a loss or discontinuation of precious research created out of hard work. This method of digital preservation will ensure that open-access journals are archived for the foreseeable future. Dark archives are private, digital archives that cannot be accessed by the general public. The data is either accessible only to a few specific individuals or completely restricted. Published journal articles are stored and preserved in dark archives as failsafe repositories (Why archiving is essential for open access (OA) journals and how to get started, 2019).

It is the place where all scholarly research is deposited securely and its content becomes public only after a “trigger event”. Dark archives will only release content when there is a “trigger event”, such as confirmation that a journal is no longer in publication. When a journal is no longer available in a publication or when a publication is discontinued, this is known as a “trigger event.” The method allows archiving OA journals for a long period of time in a way that is sustainable and long-term. When it comes to the deployment of the dark archive for the purpose of LTDP, LOCKSS, CLOCKSS and Portico are the major players with wider adoption by libraries and publishers.

## LOCKSS

LOCKSS stands for Lots of Copies Keep Stuff Safe. LOCKSS is a research-based, open-source software application providing for robust, peer-to-peer digital preservation. It is founded at Stanford University in the year 1999. It was released into production in 2004. In 2016, the LOCKSS programmes were formally integrated into Stanford Libraries Digital Library Systems and Services group. With its polling mechanism, it is able to replicate, migrate and repair files (LOCKSS Program, 2023).

As part of its digital preservation software development and support program, the LOCKSS Program develops and supports libraries using free and open-source software. Now, more than 20 networks, including libraries, government agencies and academic institutions, use LOCKSS technology to preserve content. It uses peer networks known as LOCKSS boxes for ingesting, preserving and disseminating content. Peers’ LOCKSS daemons act independently as web crawlers for collecting content and as servers or proxy servers for disseminating

it. Plugins are Java classes mostly defined in XML that provide a way to collect web content from a particular site, as well as to divide that content into Archival Units (AUs). An AU refers to a journal volume. A LOCKSS network preserves content by detecting damage to peer-to-peer content and repairing it. As a result of peer-to-peer polling, LOCKSS detects damage when it finds disagreement between instances of the same AUs on different peers and recovers damage once agreement is discovered. Peers who detect damage to their AUs request a repair from another peer who has previously agreed about the AU. The five basic functions performed by LOCKSS are:

- *Location:* In order to preserve an AU in the network, the LOCKSS box in the network must be able to locate at least some of the other LOCKSS boxes in the network.
- *Verification:* LOCKSS boxes are enabled to confirm that their content has been collected for an AU is identical to what other LOCKSS boxes preserving the same AU have collected, identifying and correcting collection errors.
- *Authorisation:* In order to repair damage, a LOCKSS box preserving an AU can provide content from the preserved AU to a second LOCKSS box preserving the same AU because the second box has proven to this one previously that it collected matching content.
- *Detection:* In the event that an AU experiences random loss or damage, the LOCKSS box may request repairs from another LOCKSS box that preserves the same AU.
- *Prevention:* The detection of non-random changes at multiple LOCKSS boxes in the network prevents attempts at deliberate modification of content in an AU.

## CLOCKSS

CLOCKSS stands for Controlled Lots of Copies Keep Stuff Safe. This is an advanced version of LOCKSS and began operating in 2010. Despite economic, technological, political and environmental disturbances,

this archiving service ensures that users have access to the content. Similar to LOCKSS, it works on the principle of replication and polling, but in CLOCKSS information is only accessible in trigger situations. CLOCKSS is a not-for-profit organisation, governed by a Board of Directors made up of 347 libraries and 496 publishers (CLOCKSS, 2022).

Using this unique service, publishers and libraries know their content will be available to those who desire access after a trigger event, despite disruptions and failures caused by technology, economics, environmental and political factors. There are seven preservation steps of the CLOCKSS system:

- *Step One:* Content presentation or source files are provided to the CLOCKSS system by the publisher. Content readers normally view presentation files that contain HTML pages. The publisher uses source files internally as minimally formatted content. A CLOCKSS-provided permission statement will tell crawlers what content is available to collect on the publisher's site, allowing CLOCKSS crawlers to access the presentation files. To allow CLOCKSS access to the publisher's source files, the publisher needs to place them on a designated FTP site.
- *Step Two:* The publisher's content is ingested into special CLOCKSS boxes located at Rice, Indiana and Stanford Universities.
- *Step Three:* It is necessary to verify that the content in CLOCKSS boxes is identical before they are allowed to use them. Content that is authoritatively sourced is established in this way.
- *Step Four:* Most of the CLOCKSS boxes serve as preservation devices, performing the main functions of storage and auditing. Preservation CLOCKSS boxes take the content from the ingest machines after it has been validated for quality.
- *Step Five:* Afterwards, the content is audited and repaired. In order to verify the preservation of content, the CLOCKSS boxes communicate continuously over the Internet. CLOCKSS boxes will receive repairs if the contents of one are damaged or incomplete by using the contents of other CLOCKSS boxes and/or the original presentation

files provided by the publisher. The CLOCKSS boxes work together so that they don't have to be backed up individually. The system also provides clear assurance that the correct content is always available and that it's functioning as expected.

- *Step Six:* When CLOCKSS Board members decide to release content from the CLOCKSS Archive after a trigger event occurs, two things happen. First, automatic migration of content to the latest format is performed, and second, the CLOCKSS boxes send their content to an organisation that hosts CLOCKSS over a public web server (Currently the EDINA Data Centre, University of Edinburgh and Stanford University).
- *Step Seven:* Stanford University and the University of Edinburgh have made the released content freely available. It is also directly available via Open URLs through Crossref, or either of the local library link-resolvers or from triggered content available on the CLOCKSS website (<https://clockss.org/triggered-content/>).

## Portico

Portico is a community-supported preservation archive that safeguards access to e-journals, e-books and digital collections. The Portico digital preservation service is part of ITHAKA, a not-for-profit organisation helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. Its unique, trusted process ensures that the researcher, scholars and students will continue to be able to access the content that is lost. Through PORTICO, content will remain accessible and usable in the future. E-journals, e-books, digitised collections and other electronic scholarly content can be accessed in Portico. It provides access to archived content only in case of a trigger event (Portico, 2023). Portico also preserves thousands of OA titles and works with publishers to maintain free access when they are triggered. Triggered journals and eBooks are accessible to all Portico participants, while triggered digital collections are available to all libraries that had previously purchased the collections; OA content is available to everyone. Portico follows the following preservation steps:

- *Preservation Planning:* As part of the preservation planning process, Portico analyses the formats and packaging of the content, as well as developing any tools that might be required.
- *Receipt and Inventory Management:* As soon as content is provided by the provider, it is transferred to Portico.
- *Processing and Archival Deposit:* During the ingest process, Portico validates files, extracts relevant technical and descriptive metadata and records all actions taken on the content.
- *Monitoring and Engagement:* In order to ensure security and accessibility of the archive, Portico manages the content and archive on an on-going basis.
- *Content Delivery:* Participating institutions are allowed to access content under certain conditions.

## Related Studies

Ubaid Ullah and Gul, Sumeer (2019) opined that the preservation of digital material was more complicated than print material. Libraries have reduced a great deal of stress in preserving digital content and information by adopting digital preservation policies and initiatives, including LOCKSS, CLOCKSS, PORTICO and many more, which have been implemented in light of ever-changing technology, making it safe and accessible for future generations as well. Mering, Margaret (2016) explained five initiatives for the preservation of digital resources, including LOCKSS, CLOCKSS and Portico. The author articulated that we should rely on these initiatives and that they were very important for the future. In addition to preserving electronic scholarship, they ensure that it is accessible forever. This opens the door for scholarships to increasingly rely on electronic content and have confidence in its stability and accessibility. Kiefer (2015) explored the development of digital preservation, the challenges that preservation of multiple formats of digital scholarship brings with major focus on the CLOCKSS Archive. The author articulated that the CLOCKSS Archive continues to evolve in many different ways to meet the challenges of long-term preservation. Shah, Mering and Margaret (2015)

explained in their research that there was a collaboration between libraries, non-profit organisations and the publishing industry that led to LOCKSS, CLOCKSS, PORTICO and other digital preservation initiatives. By preserving and making permanently available electronic scholarship, they envision a future in which all scholars are able to rely on electronic content without fear of its stability or availability. Seadle, Michael (2011) described in his study that only 8% of the directory of open access journals (DOAJ) titles are in LOCKSS/CLOCKSS and only 5% in Portico. The findings also suggest that it could take eight years to archive all full-text directory of open access journals (DOAJ) articles in e-Depot based on current plans. The most important implication was observed: the most OA titles listed in directory of open access journals (DOAJ) currently have no effective long-term digital archiving. Reich, Victoria (2009) explained how LOCKSS, CLOCKSS and Portico work. This article studied the trigger event that occurred in Sage Publications from 2001 to 2008 in OA Journals. It showed the triggered content in different years from 2001 to 2008.

## Study Objectives

The research objectives of the study are to:

- To examine the dark archiving policies of OA journals indexed by DOAJ.
- To study the status of the adoption of dark archive platforms - LOCKSS, CLOCKSS and Portico-with regard to OA journals.
- To examine the preservation status and trigger events pertaining to dark archives powered by LOCKSS, CLOCKSS and Portico.
- To examine the status of supporting libraries and participating publishers in context with the dark archives.

## Research Design

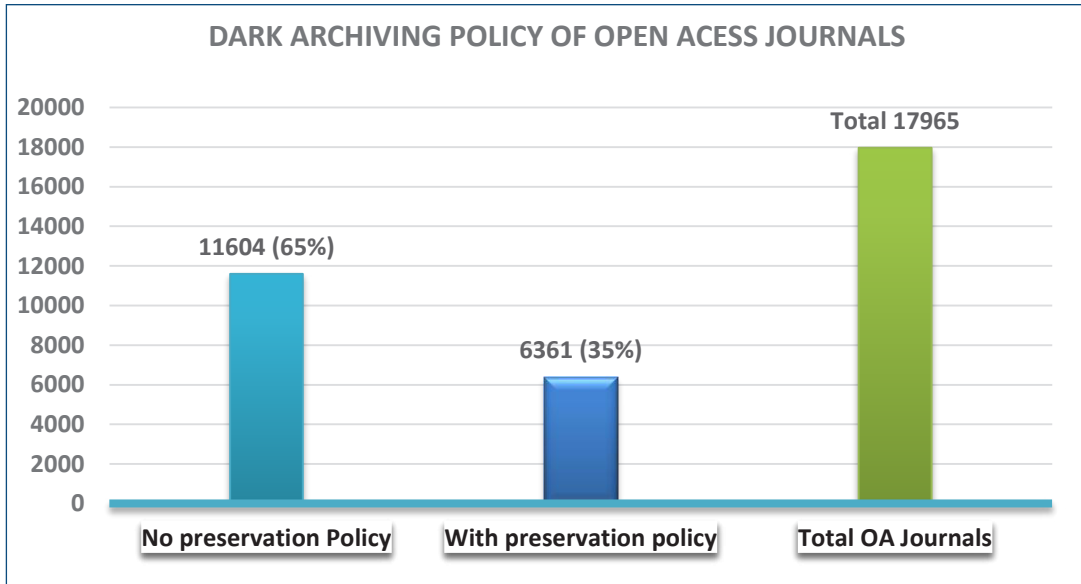
To accomplish the objectives of the study, the data was extracted from the DOAJ). DOAJ is a community-created

online directory that indexes and provides access to high-quality, open-access, peer-reviewed journals. It contains quality and authentic data regarding 17978 open-access journals that are being currently indexed by them. A total of 17965 (99.93%) records regarding the adoption and implementation of long-term preservation data were extracted as a CSV file. Data cleaning was performed to remove the errors and anomalies using the web tool Data list (<https://app.datablist.com/>). The final data set was analysed using spreadsheet software for obtaining results and findings. The data regarding preservation status and trigger events associated with the dark archives and the status of supporting libraries and participating publishers in context with the dark archives were obtained from their respective websites.

## Results and Discussion

### Dark Archive Policy for Open Access Journals

Fig. 1 shows the dark archiving policy of OA journals. As the below Fig. 1 displays, only 35% of open-access journals use a dark archive policy, whereas 65% of the journals do not have any dark archive policy. OA Journals aims to provide OA to scientific literature for instant dissemination of the latest appraises in science and technology by establishing a reliable platform. They serve the society by making their content available freely online, thereby alleviating the access challenges presented by subscription journals. They are beneficial to almost all sectors of society, whether it is science & technology, education, health care, or the education sector. Hence, it is really important to preserve these resources through a proper preservation policy. It is a matter of concern for the journals indexed by DOAJ without having any LTDP policy, such as a dark archive, posing uncertainty over their future availability. It is recommended that the publishers associated with these 35% OA journals should take timely, effective and sustainable initiatives to safeguard the valuable public good for society.



**Fig. 1: Dark Archiving Policies for Open-Access Journals**

• *Adoption of LOCKSS, CLOCKSS and Portico*

OA journals indexed by DOAJ have several long-term preservation policies, including LOCKSS, CLOCKSS, etc. Table 1 below depicts the adoption status of LOCKSS, CLOCKSS and PORTICO web tools as the dark archive-enabled preservation policy.

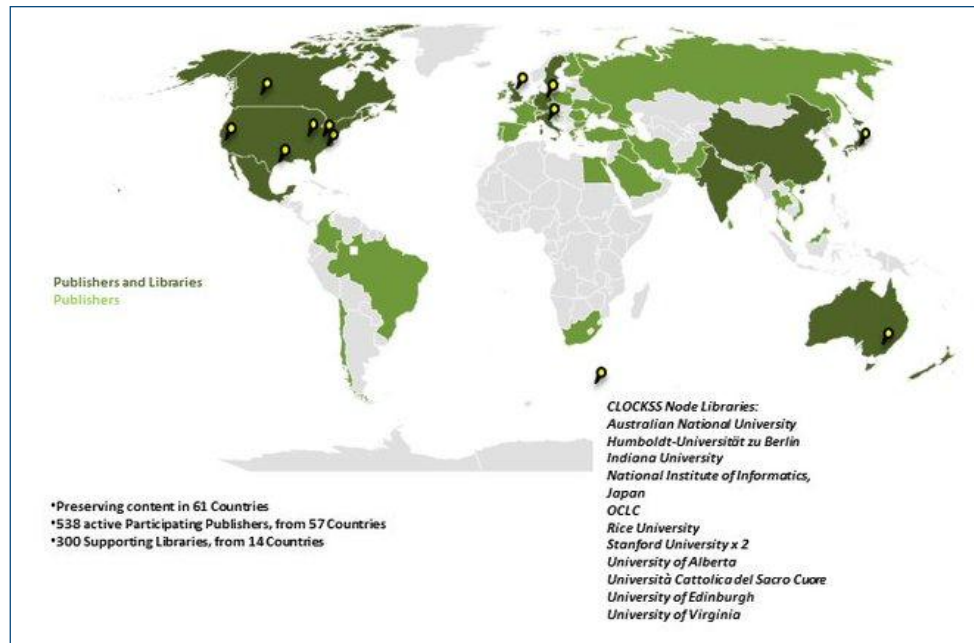
**Table 1: Adoption Status**

Dark Archive	Preservation Service Instances
LOCKSS	1828 (23%)
CLOCKSS	3191 (39%)
Portico	3098 (38%)
<b>Total</b>	<b>8117 (100%)</b>

The table above presents the number of preservation service instances in open-access journals. It is observed that CLOCKSS has the maximum preservation instances (39%) of instances, followed by Portico (38%) and then LOCKSS (23%). CLOCKSS runs on LOCKSS technology, although there is a slight difference.

The LOCKSS (developed by Stanford University) name has come over time to refer to at least 3 different things:

- The LOCKSS technology is “fault tolerant” and safeguards against the long-term, well-documented causes of digital loss: human error, computer attacks and economic and organisational failure.
- The LOCKSS technology preserves all formats (video, sound, pictures, text, etc.) and genres of content (e-journals, e-books, conference proceedings, etc.).
- It is in the name of the Global LOCKSS network, which provides distributed preservation and local post-cancellation and perpetual access for subscription electronic journals and books, as well as a mechanism for building local collections of web-based scholarly OA publications.
- A community of varied distributed preservation services all operating on LOCKSS software. These preservation services are the Alabama Digital Preservation Network (ADPN), Cariniana, CLOCKSS, MetaArchive, the Michigan Digital Preservation Network (MDPN), SAFE PLN, US Docs PLN, CGI PLN and PKP PLN.



*CLOCKSS is currently supported by 538 publishers in 57 countries (light green) and 300 libraries in 14 countries (dark green).*

CLOCKSS is a global archive that preserves content on behalf of all libraries and scholars worldwide. CLOCKSS does so in 12 strategically chosen academic libraries across the globe to optimise the content’s safety against political and environmental threats. It is a shared dark archive that runs on LOCKSS technology. CLOCKSS’s content is hosted on 12 servers (CLOCKSS Archive nodes) around the world, at leading academic libraries, with robust infrastructure and security. Located in Asia-Pacific, Europe and North America, they each have a complete copy of the archive. They perform the preservation service on behalf of libraries and users everywhere; other libraries do not need to host the published scholarly literature for preservation purposes themselves. Many libraries support CLOCKSS through financial contributions and by participating in Archive governance; they do not need to host any technology locally. CLOCKSS is considered a “controlled LOCKSS” archive because of the unique role of the 12 nodes. Through CLOCKSS and Portico, everyone can view the triggered content in open-access journals, but in LOCKSS only the participating institutions can view the triggered content. It could be a valid reason for the higher adoption

of CLOCKSS and Portico as sustainable LTDP measures over LOCKSS. It was observed that multiple preservation strategies involving dark archives are being followed by a single journal or publisher, i.e., a single journal might be using LOCKSS and CLOCKSS as well as Portico.

- *Preservation Status and Trigger Events for CLOCKSS and Portico*

**Table 2: Preserve Content by CLOCKSS**

<i>Preservation Status and Trigger Events</i>	<i>Number of Records</i>
Preserved Journal Titles	35,000
Preserved Books	400,000
Preserved Journal Articles	51 million
Triggered Titles (Open Access)	66
Triggered Articles	13,000
Triggered Books	0

Trigger events happen when content is permanently removed. Such cases include:

- *Publisher No Longer in Business:* The publisher is no longer in business or is no longer in the business of publishing content or providing access to previously published content, and there are no successor interests or reversions or transfers of rights.

- *Title No Longer Offered:* The publisher is no longer providing access to the content, and there are no successor interests or reversion or transfer of rights.
- *Back Issues No Longer Available:* The publisher has stopped offering or providing access to some or all of the back issues of the content and there are no successor interests or reversion or transfer of rights.
- *Catastrophic Failure:* While still publishing content, the publisher is not able to provide access to the content electronically due to technical or similar catastrophic and permanent failure.

The materials are made available to everyone for free and OA in case content is affected by a trigger event.

Table 2 presents the status of preserved content using CLOCKSS. If the content that is held in the Archive disappears (or is about to disappear) from the Web, CLOCKSS will “trigger” it for Open Access. Since the launch of CLOCKSS in 2008, 13,000 articles from 66 journals have been triggered – a very small fraction of the 35,000 journal titles that the Archive preserves-and it also been observed that no such triggered event occurred in the case of books.

**Table 3: Preserved Content by Portico**

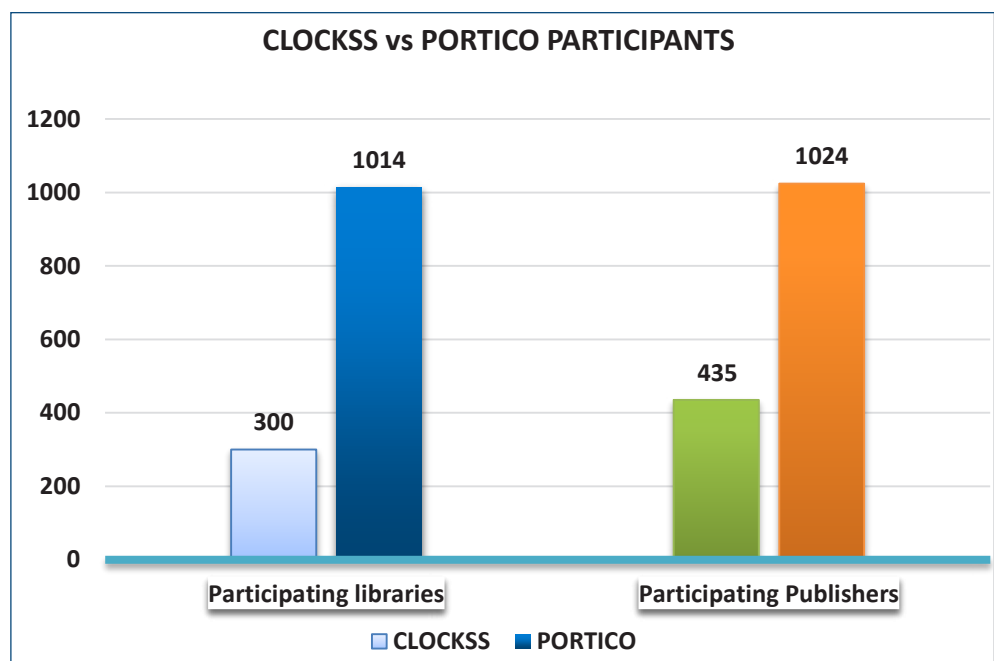
<i>Preserved Items</i>	<i>No. of Records</i>
Preserved archival units (articles, books, etc.)	137,860,628
Preserved e-journal titles	36,352 out of 38,834 e-journals committed
Preserved ebook titles	1,602,380 out of 2,010,899 e-books committed
Preserved files	2,300,648,691
Preserved images	1,359,912,004
Preserved repository-created archival files	289,109,165
Preserved supplied text files	484,998,737
Preserved application-specific files	32,309,665
Preserved multi-file packages	17,286,605
Preserved video files	241,686
Preserved audio files	8,367
Preserved executable files	100
Triggered events	207

Portico is one of the most comprehensive digital preservation archives; holding more content than others and having extensive knowledge about the content reserve, including an understanding of component file types and formats. The archive is gaining acceptance and updated regularly. Table 3 depicts the preserved content and triggered events by Portico. It has a variety of content being preserved including books, videos, audio, images, etc. It has the highest number of preserved, contents as

compared to CLOCKSS. It can be observed from Table 3. Those 207 triggered events occurred in Portico.

- *Participants for CLOCKSS and Portico*

Fig. 2 shows the number of participants in CLOCKSS and Portico. It can be depicted from Fig. 2 that Portico has almost 3 times more participants than CLOCKSS. Similarly, the number of publishers is also higher in Portico than CLOCKSS.



**Fig. 2: Participants in CLOCKSS and PORTICO**

Table 4 displays the top 10 countries whose libraries and publishers participate in Portico. It can be seen that Brazil has the highest number of libraries, and the United States has the highest number of participating publishers

in Portico. On the other hand, Australia has the lowest number of participating libraries, and South Africa has the lowest number of participating publishers among the top 10 participating countries.

**Table 4: Participation by Countries (Portico)**

Sr. No.	Participating Libraries		Participating Publishers	
	Country	Number	Country	Number
1	Brazil	420	United States	353
2	United States	397	United Kingdom	109
3	Portugal	80	Italy	65
4	Germany	66	India	37
5	Argentina	59	Canada	33
6	Greece	54	Colombia	26
7	United Kingdom	54	Germany	25
8	Italy	37	Australia	25
9	Switzerland	26	Spain	22
10	Australia	25	South Africa	21
11	Others	93	Other	316
	Total	1,311	Total	1,032

## Discussion and Recommendations

Digital objects have no meaning on their own. They are not static and reified; they do not remind us of their value by sitting on a shelf within reach with their titles in view on their spines or on their covers. The threats that put this information at risk also exist out of our sight and could make these digital objects irretrievable. A comprehensive means of preserving, safeguarding and making accessible digital objects for the future, especially those objects that comprise the record of advances in human knowledge, is an essential foundation for human progress. That is what digital preservation is all about. Digital preservation can never be a solved problem. It is work that does not finish and it becomes harder over time as formats, software and hardware fade into memory and the creators and publishers move on to new challenges.

Preservation requires active management to ensure that content and data are and remain, healthy. Digitised content, backup copies and content in storage can deteriorate over time.

Research libraries join CLOCKSS to ensure the world's digital content is protected for future generations of researchers. Publishers contribute the content they publish to CLOCKSS to preserve it for the long-term in a secure digital archive without incurring the expense of building their own preservation and storage system. When a publisher's content is triggered, libraries and researchers are not charged for access; instead, the triggered content is made available OA and for free. Digitised content must be protected and conserved under strategies, programmes and policies. The adoption of dark archives is extremely low when compared with the quantum of published research, raising question mark on its long-term preservation. There are a few recommendations for institutions, libraries and publishers for the adoption of dark archives for the preservation of digital resources.

- *Policy*: There is a need for a well-planned required policy for the libraries, archives, museums and publishers for integration with dark archives. An organisation or library can easily follow a policy as long as it is comprehensive, written and approved. 65% of open-access journals have no preservation policy. There should be mandates for preserving the OA research. Strong policies should be framed for

the implementation of dark archives for publishers and libraries in each country. Such initiatives should be applicable for both subscriptions-based as-well-as OA literature.

- *Create Awareness*: There is a need to raise awareness of the importance of dark archives as a sustainable long-term preservation platform. There are thousands of journals published all over the world, but publishers and libraries are not much aware and concerned about their preservation through digital archives. Libraries and publishers should be motivated to participate in CLOCKSS and Portico.
- *Coordination*: There should be healthy coordination among publishers and libraries. Publishers and libraries need to recognise their shared responsibility to perpetuate the distribution of knowledge and preserve the content for researchers, academics and students.
- *Central Archiving Body*: There should be some central archiving body in every country which facilitates provisions of Dark Archives.
- *Integration*: Mechanisms for interoperability of conventional institutional repositories and digital libraries with dark archives to be ensured.

Dark Archives like CLOCKSS exist to preserve the hard work and knowledge of scholars by ensuring that archiving remains in the hands of the community. By contributing to CLOCKSS, organisations become part of that community of libraries and publishers, and they gain a direct voice in archiving decisions. The active support of libraries strengthens the influence of dark archives with publishers, enabling them to persuade more to entrust the content they publish to the archive. Moreover, dark archives facilitate libraries engaged in publishing to preserve their important publications to become part of the international scholarly record preserved in CLOCKSS. CLOCKSS complements and supports libraries' own digital preservation strategy. It is more cost-effective and efficient to collaborate to preserve published digital books and journals rather than each institution duplicating this effort. CLOCKSS also saves library money by reducing the risk. Publishers who participate in CLOCKSS agree that they can make triggered content available open access and grant them the rights to do this in perpetuity.

Library support enables us to subsidise preservation costs for the long-tail of new and smaller publishers whose content is most at-risk. For example, this includes diamond OA titles via the JASPER Project (FAQ - CLOCKSS, 2023).

## Conclusion

In the context of the popularity in terms of adoption of dark archives as LTDP and digitally preserved contents, it can be inferred that publishers and libraries are either using CLOCKSS (39%) or Portico (38%) for digital preservation, while the adoption percentage for LOCKSS was found to be 23%. But overall, there exists a severe lack of wider acceptance of dark archives. This is also true if usage of dark archives is seen in a geographical context, as the phenomenon is not globally visible.

The study concluded that only 6,361 (35%) OA journal titles have a long term digital preservation (LTDP) policy. As knowledge production has progressed, preservation of the digital resources has become increasingly important. The publishing industry is also shifting from print to digital publishing with the rise of the subscription-based model of publishing. Journals are, therefore, not always guaranteed to be accessible in the long term and they may even disappear entirely from the web if their authors discontinue publication. There is always the risk of electronic resources being compromised by a variety of threats, including hardware failure, software failure, natural disasters and economic downturns. During the transition from analogue to digital, there was uncertainty as to who was responsible for preserving scholarly literature in electronic formats: publishers, libraries, authors, or academic institutions. Without a general agreement concerning who is responsible for preserving electronic resources, no one will be accountable and we could lose large parts of the scientific record if we do nothing.

Through digital preservation policies, it is possible now to archive digital information perpetually, though it is more difficult to preserve the digital collection than the print collection. It is essential to have policies in place to ensure that these digital records remain accessible, trustworthy

and verifiable. The dark archive is a long-term digital preservation (LTDP) policy enabling libraries to mitigate risks and threats to valuable knowledge accumulated over centuries. Dark archive initiatives including LOCKSS, CLOCKSS and PORTICO can play an instrumental role in ensuring risk-free perpetual access through need-based triggered events. Implementing digital preservation policies through such initiatives can make content and information safe and accessible to future generations as well as the present in spite of the ever-changing technology. The rising acceptance and deployment of dark archives by the libraries and publishers is important for the artefacts signifying research and cultural heritage. These innovative approaches enabled by technological advancement must be adopted by more and more libraries, archives and museums (GLAMs) to facilitate knowledge when needed. Knowledge once created must be preserved for posterity so as to serve as the building block to the knowledge society, irrespective of the challenges posing a threat to their existence.

## References

- Burnhill, P., & Guy, F. (2010). Piloting an e-journals preservation registry service (PEPRS). *The Serials Librarian*, 58(1-4), 117-126. doi:<https://doi.org/10.1080/03615261003622742>
- CLOCKSS. (n.d.). Retrieved from <https://clockss.org/>
- Concept of Digital Preservation. (2019, March 28). UNESCO. Retrieved from <https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation>
- FAQ - CLOCKSS. (2023, January 11). CLOCKSS. Retrieved from <https://clockss.org/faq/>
- Katharina, M. (2022, March 22). Would you rather use LOCKSS, CLOCKSS or Portico? Three approaches to long-term preservation of scholarship. *IFLA Section Academic and Research Libraries* (blog). Retrieved from <https://blogs.ifla.org/ar/2022/03/22/would-you-rather-use-lockss-clockss-or-portico-three-approaches-to-long-term-preservation-of-scholarship/>
- Kiefer, R. S. (2015). Digital preservation of scholarly content, focusing on the example of the CLOCKSS archive. *Insights the UKSG Journal*, 28(1), 91-96. doi:<https://doi.org/10.1629/uksg.215>

- Laakso, M., Matthias, L., & Jahn, N. (2021, February). Open is not forever: A study of vanished open access journals. *Journal of the Association for Information Science and Technology*. doi:<https://doi.org/10.1002/asi.24460>
- Malak, H. A. (2022). *What is digital preservation? Why is it important? Information management simplified*. Retrieved from <https://theecmconsultant.com/what-is-digital-preservation/>
- Open Access Journals. (n.d.). Retrieved from [www.openaccessjournals.com](http://www.openaccessjournals.com).
- Reich, V. (2009). From dark archive to open access: CLOCKSS trigger event lessons. *Against the Grain*, 21(2). doi:<https://doi.org/10.7771/2380-176x.2552>
- Seadle, M. (2011). Archiving in the networked world: Open access journals. *Library Hi Tech*, 29(2), 394–404. doi:<https://doi.org/10.1108/07378831111138251>
- Staff, Webopedia. (2004, April 13). *What is dark archive?* Retrieved from <https://www.webopedia.com/definitions/dark-archive/>
- Why Archiving is Essential for Open Access Journals and How to Get Started. (2019). Retrieved from <https://blog.scholasticahq.com/post/why-archiving-essential-for-open-access-journals-how-to-get-started/#:~:text=The%20best%20way%20to%20ensure,raise%20awareness%20of%20published%20articles>