



LOCKSS: Re-establishing Librarians as custodians of journal content

Over the last decade libraries have increasingly shifted journal access from print to digital. The preference of users for online content, the demand of readers for a broader range of content, and the rising costs of library shelf space all contributed to this change. However, current publisher distribution models require library users to access content hosted on a centralised publisher-maintained web server. The LOCKSS (for Lots of Copies Keep Stuff Safe [1], [8]) approach helps libraries re-gain custody of journal assets while maintaining the access and license restrictions stipulated by the publisher.

The changing model of access

Web publishing provided ease of access, faster dissemination and cost and space benefits to libraries. The development of this model led to two significant by-products. Publishers wished to restrict access to users who had subscriptions, most easily achieved by requiring authentication to a central server through methods such as Athens and IP range restrictions. There are many licensing options under which content can be accessed: annual subscriptions, bulk basket deals, short term backfile access, and aggregators are all widely used and variations in licenses apply to each option. Dynamically updating the access and usage conditions was reasonably easily achieved through a single resource.

Journal users and librarians were concerned that this centralised model was characterised by a single point of failure. It lacked the redundancy inherent in the physical model, where the number of copies distributed among libraries provided some measure of security against loss of access. In this digital environment what happens in the event a publisher disappears or a journal changes hands? How can libraries minimise the risk to these fragile digital materials? To address this a number of journal archiving approaches have been established ([7], [10]). Legal deposit of published content provides some base assurances but is limited to material relevant to the applicable country and access may be heavily restricted (often either to on-site access or a very small number of concurrent copies at one time). Third party non-profit archiving services are under development ([2], [3]) and offer some guarantees of access in the event a title is no longer available. The LOCKSS approach establishes individual archives within each participating library, facilitating the development of a persistent, well managed collection of content relevant to the libraries objectives.

The LOCKSS System

The LOCKSS system, under development at Stanford University Libraries since 1999, is open source software allowing libraries to collect, maintain, and access local copies of journals meaning they own rather than lease this content. The LOCKSS software is most commonly made available as a “platform CD” used to turn a computer into a dedicated appliance. The platform CD significantly automates the configuration of a LOCKSS system. A system administrator must identify a set of prerequisite network properties, but all other installation options are standardised and preconfigured. The LOCKSS software is never installed on the hard disk in the traditional sense. The only thing that persists on disk across reboots is the content the box collects and preserves. Each time the machine is booted, the operating system is reinstalled from the CD into a memory filesystem backed by the swap partition. Written entirely in Java, the LOCKSS software is freely distributable and, in keeping with the terms of the BSD license, the source code is open and freely modifiable by programmers on the condition that modifications are also free and open. The standalone LOCKSS daemon is intended to run on any UNIX-like system.

A number of drivers have influenced the development of the LOCKSS software ([4], [9]), intended to ensure the system is sustainable over an extended time frame.

- Media, hardware, software failure and obsolescence: To minimise the risk of hardware and software obsolescence the LOCKSS system operates across a diverse set of hardware and software platforms. The distributed environment removes reliance on centralisation, minimising the risks of hardware failure. The self-repairing LOCKSS audit protocol removes the need for costly backup policies.
- Communication Error and Network Failure: Network communications can be unreliable. Network communications can be unreliable. In order to correct data errors that may occur during ingest or migration across hardware, the LOCKSS audit protocol automates their discovery and repair.
- Operator Error, or External and Internal Attack: Well meaning but inexperienced system operators can unintentionally damage data or leave a system vulnerable. By limiting system administration during installation and archive administration during daily operation, accidental errors are avoided. Similarly, minimising the range of installed software applications, and restricting both physical and virtual access to the machine counter the potential for external or internal attack.
- Economic and Organisational Failure: Library budgets fluctuate and sustained investment cannot be guaranteed. Sharing the responsibility for the intellectual content of electronic journals through distributed custody of the data avoids a single point of failure.

Content Collection and Access

When archiving content, the LOCKSS system harvests an identical copy of journal content from a publisher website which is then stored in the local LOCKSS box. Content is collected in manageable units, often corresponding to subscription units used by publishers. These are termed Archival Units (AUs). Typically, a single archival unit collected by the LOCKSS system matches a complete journal volume.

For each publishing platform, a LOCKSS plugin is used to ensure the LOCKSS daemon correctly collects only the material relevant to a single archival unit. A series of rules corresponding to URLs, known as regular expressions, are used to define the collection of content comprising an archival unit. By harvesting from the publisher's website, the LOCKSS system collects a particular rendition rather than source files. The result is that when users access the archived material the look and feel will match that intended by the original publisher. The LOCKSS system is format agnostic; any format that can be transmitted over the web can be archived. This is especially relevant in the emerging journal sector as supplementary datasets and innovative presentation formats proliferate alongside the more traditional text-based journal format.

The LOCKSS system contains a small web server through which content is made available. By design, the LOCKSS software was intended to be integrated with an institutional proxy, meaning access to content is transparent to users: there would be no interruption of service between content accessed from the original location to that from the LOCKSS box. In response to library feedback, LOCKSS is developing the components that will allow preserved content to be discovered through library link resolver systems.

The LOCKSS Alliance

The LOCKSS Alliance [6], established in 2005, is a membership organisation of more than 90 libraries intended to offer institutions a forum to share experiences and concerns related to LOCKSS and journal archiving more generally. Members are also offered strategic opportunities to help determine long-term priorities and directions for the evolution of the LOCKSS software and program. Membership requires an annual fee, with the cost based upon the institution size and budget. Although open access journal content is available to all, Alliance membership gives access to premium subscription-based content, ongoing support, and direct engagement with the LOCKSS development team.

References

- [1] LOCKSS. www.lockss.org
- [2] Portico. www.portico.org
- [3] CLOCKSS. www.clockss.org
- [4] David S. H. Rosenthal, Thomas S. Robertson, Tom Lipkis, Vicky Reich, Seth Morabito, "Requirements for Digital Preservation Systems: A Bottom-Up Approach", D-Lib Magazine, Volume 11 Number 11 November 2005
www.dlib.org/dlib/november05/rosenthal/11rosenthal.html
- [5] OpenURL.
www.exlibrisgroup.com/sfx_openurl.htm
- [6] LOCKSS Alliance.
www.lockss.org/lockss/LOCKSS_Alliance
- [7] Anne R. Kenney, Richard Entlich, Peter B. Hirtle, Nancy Y. McGovern, and Ellie L. Buckley, "E-Journal Archiving Metes and Bounds: A Survey of the Landscape", CLIR Report, September 2006. 120 pp.
www.clir.org/pubs/abstract/pub138abst.html
- [8] Vicky Reich, "Editors' Interview with Victoria Reich, Director, LOCKSS Program", Volume 10, Number 1, Feb 15 2006.
www.rlg.org/en/page.php?Page_ID=20894
- [9] David S. H. Rosenthal, Thomas Lipkis, Thomas Robertson, Seth Morabito, "Transparent Format Migration of Preserved Web Content", D-Lib Magazine, Volume 11, Number 1, January, 2005
www.dlib.org/dlib/january05/rosenthal/01rosenthal.html
- [10] Maggie Jones, "Review and Analysis of the CLIR Report E-Journal Archiving Metes and Bounds: A Survey of the Landscape", February 2007.
www.jisc.ac.uk/whatwedo/programmes/programme_preservation/ejournalarchiving.aspx

Key Facts about LOCKSS

The LOCKSS approach has been commended by both libraries and publishers for providing the academic community with a robust, cost-effective tool to take responsibility for preserving e-journal content. The LOCKSS Alliance has more than 125 participating libraries. With another 60+ libraries running LOCKSS but not currently participating in the Alliance, more than 185 libraries are involved in LOCKSS. The full list of participating libraries can be found at: <http://www.lockss.org/lockss/Libraries>.

Now in its tenth year, LOCKSS runs on open source software which won the 2004 Association for Computing Machinery's Peer Reviewed Award for Best Research. Over 300 publishers have committed journal content for preservation in the LOCKSS system. The complete list of publishers and titles can be found at: http://lockss.org/lockss/Publishers_and_Titles.

The JISC/CURL funded UK LOCKSS Pilot Programme is a leading initiative to establish a national LOCKSS Alliance and collectively preserve journal content relevant to UK priorities. The activity was led by a dedicated support service embedded within the UK's Digital Curation Centre, and brought JISC and their negotiation agent, Content Complete, together with a diverse set of thirty Higher Education libraries. Library budgets were spread across the JISC banding arrangements and e-journals accessible from these institutions varied in quantity (from approximately 7,500 to over 20,000 titles), made available through many combinations of publishers and aggregators. The UK LOCKSS Pilot Programme is discussed in detail at: <http://www.ijdc.net/ijdc/article/view/49/55>.

Conclusions

In response to the concerns of librarians, publishers are increasingly participating in efforts to develop effective journal archiving strategies. The LOCKSS system provides a critical component in the journal distribution infrastructure, allowing libraries to take custody of assets for which they have paid, while conforming to the licensing arrangements they have agreed with publishers.

The LOCKSS approach takes steps to ensure libraries are responsible not only for short-term access, but involved at many stages in the emerging model of journal archiving. It is certainly the case that the technologies and licensing agreements will continue to develop and evolve, ensuring that both libraries and publishers acquire, secure, and maintain the rights, access conditions, and financial benefits that they both expect and deserve.