A Persistent Access Preservation Program
Answers for Library Directors

What problem does the LOCKSS program address?

The collections of the world's libraries hold much of society's memory. The Web is disrupting the ability of libraries to fulfill this important role for society. Librarians have lacked the tools to build collections of digital material, and to preserve the material they have managed to collect against the tendency of bits to decay so that they remain accessible to future generations. A library without a collection loses its raison d'etre.

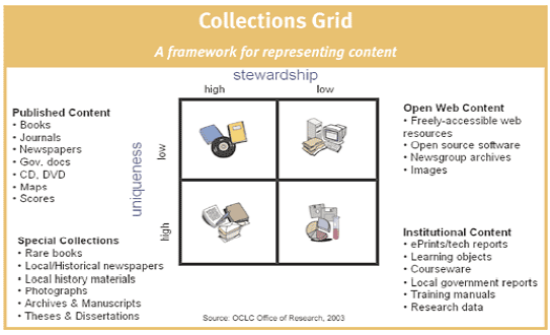
These problems are exacerbated by the rigorous copyright protection afforded to digital content. Preserving digital material requires explicit permission from the copyright owner. Obtaining that permission can be so onerous that libraries may only attempt to collect and preserve material their own institution publishes.

The LOCKSS program has addressed these issues of custodial responsibility, preservation and permission by a compromise between the interests of publishers and librarians. In the LOCKSS system, publishers grant libraries the right to collect and preserve access to their own copy of copyright material from the Web, just as they would if it were on paper. But the access restrictions under which the library originally obtained the material from the publisher continue to apply to the preserved copy, reassuring the publisher that the preserved copies will not be used to undermine their business.

The LOCKSS system makes collecting and preserving access to Web-based information cheap and easy. It is an open-source, peer-to-peer persistent access preservation system, which has been under development since 1998 and is deployed at almost 100 libraries worldwide. It collects content using a web crawler, just as search engines do, and disseminates it using a web proxy, just as web caches do. Access to preserved content is transparent; if for any reason a library's reader cannot access the publisher's copy, the library's preserved copy satisfies the request. This happens whether the publisher's copy is inaccessible because the subscription has been cancelled, the publisher has gone out of business, or there is a temporary network outage.

How does the LOCKSS program enable collection management and how does it relate to IRs?

Institutional Repositories (IRs) are designed for content whose copyright is held by the institution. Depositing material from other publishers typically requires time-consuming case-by-case negotiations.
Publishers grant a blanket license allowing any authorized library to use the LOCKSS system to collect and preserve a copy. A single negotiation by a single library is all that is needed. The LOCKSS program is thus the only initiative that makes it economically and technically feasible for libraries to take custody of, and to ensure long-term access to Web content they currently lease (subscriptions) or access (open-access) from other publishers. Some institutions are using the LOCKSS software to implement an IR. The LOCKSS system is designed to manage the content on the left side of this grid. IRs are designed to manage the right hand side of the grid.



Does every library need to collect all the materials they might want to access in the future?

Yes, a library's LOCKSS boxes must collect everything that the library may wish to serve to current and future readers.

Early Mellon grants showed publishers were unwilling to leave future access rights in the hands of "third parties" unless access rights were very constrained (cf. the agreement under which Elsevier deposits content at the KB). By enforcing the rule that each library have its own copy of the content and use it only for its own access, the LOCKSS system eliminates "third parties" and can leverage the existing relationship between publisher and library to allow a single negotiation of a blanket license. A terabyte of storage (10,000 years of a large journal) costs about as much as an hour of a lawyer's time so trading more storage for fewer lawyers makes sense.

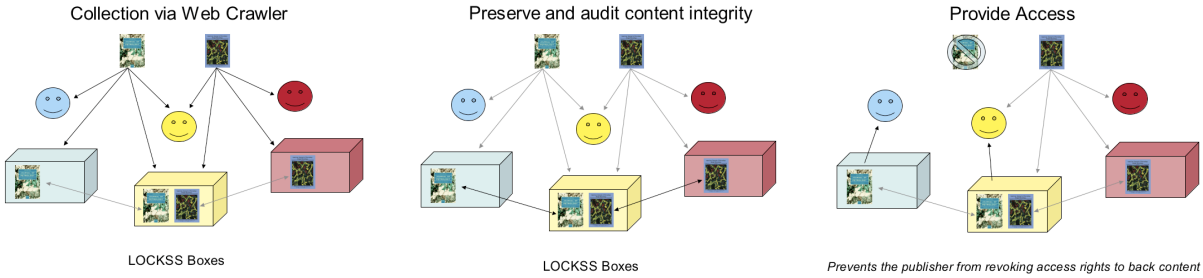
Publishers are concerned about libraries and others "stealing their content". They need to control who collects their content and monitor how it is used. Thus, content is not shared among LOCKSS boxes.

Readers at institutions that did not collect it from the publisher do not get access via an institution that did. The normal subscription agreement applies to an institution's use of the LOCKSS system (only authorized subscribers get access plus, perhaps, course packs, ILL, etc.); permission to collect and preserve content is an addition to it, not a separate agreement. One library negotiates this addition to the subscription agreement, which then applies to all libraries using the LOCKSS system. Permission to have the content is bundled with the content, simplifying the tracking of rights into the future.

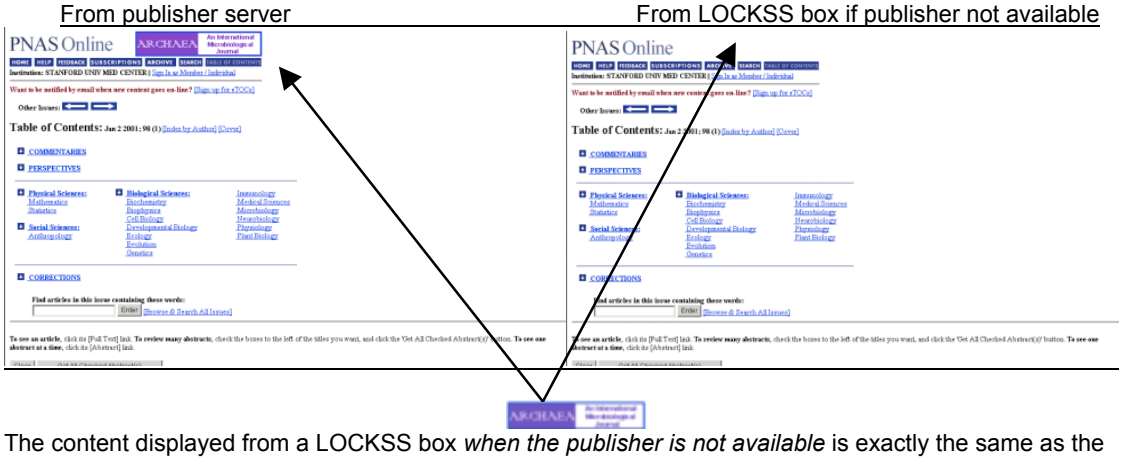
Libraries collect subscription materials at the time of publication. Most subscription publishers limit LOCKSS collection to current subscribers and currently published materials. For example, take a journal year starting in January 2006 and ending in December 2006. The publisher will grant permission to collect this material in January 2006, and revoke permission to collect it sometime in January 2007. Libraries that don't participate in 2006 can't preserve the 2006 content. This limitation on collection limits the risk to the publishers and has been key to obtaining their support; it also motivates libraries to participate sooner rather than later.

How does it work?

Each LOCKSS box is an inexpensive PC running free software. It collects crawls publisher's Web sites to collect content, takes part with other LOCKSS boxes that have collected the same content in a peer-to-peer process of mutual audit and repair, and disseminates the content by acting as a Web proxy in the same way that institutions use Web proxies (e.g. EZproxy) to provide off-campus readers access to subscription materials. LOCKSS boxes collect, preserve and disseminate all Web file formats (text, sound, pictures, video, spreadsheets, etc.). The result is that a library's readers see the material at its original URL even after the publisher has removed it, the library's subscription has lapsed, or the network to the publisher has failed.



What does the content delivered from a LOCKSS box look like?



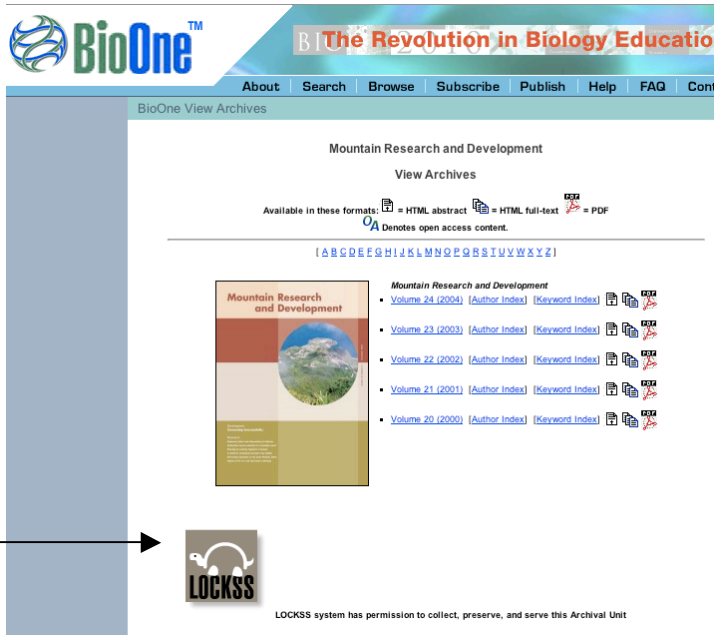
The content displayed from a LOCKSS box *when the publisher is not available* is exactly the same as the content displayed from the publisher *with one exception.* Content that changes when a reader hits their brower's reload button is not collected by the LOCKSS system. In this example, the ad changes and therefore is not collected or preserved. This dynamic content, content that has no authoritative version, is not managed by any existing digital preservation system at this time.

How do publishers give permission?

Publishers must give libraries permission to collect, preserve, and provide access to their content. In the LOCKSS system permission is granted to all libraries and is not negotiated institution by institution. The suggested license language permits libraries to:

- Collect and preserve currently accessible materials;
- Use material consistent with original license terms;
- Provide copies to other LOCKSS boxes only for purposes of audit and repair.

Permission is granted to the LOCKSS software through a web page called a LOCKSS publisher manifest. This page must contain either a suitable Creative Commons license, or the following text: "**LOCKSS system has permission to collect, preserve, and serve this Archival Unit**." An "Archival Unit" is typically a volume of a journal, and is defined by the top-level URLs included in the manifest page.



What else can I do with my LOCKSS box?

Libraries are beginning to use the LOCKSS software to capture, preserve, and give access to a wide variety of *web-based* content beyond subscription and open access electronic journals:

- **Web sites and Special Collections** As one example, the LC NDIIPP is funding Emory and five partner institutions to collect and preserve at risk digital content (e.g. online exhibitions, digital masters, cultural history Web displays). Using a "LOCKSS private network" as the technology infrastructure, the group is defining workflow processes and best practices for cooperative preservation http://www.metaarchive.org/index.html "LOCKSS private networks" are also being installed within single institutions to preserve similar materials.
- **Electronic thesis and dissertations**: Eight leading ASERL libraries are using the LOCKSS technology and OAI-PMH to collect and preserve electronic theses and dissertations. http://www.solinet.net/resources/resources_templ.cfm?doc_id=3680
- **Government Documents:** In Spring, 2005 the GPO announced a pilot project to use the LOCKSS technology to evolve the Federal Depository Library Program (FDLP) and the International Exchange Service (IES) from print to online. More than 20 participating libraries have a substantial voice in determining the future of these programs. http://www.access.gpo.gov/su_docs/fdlp/lockss/

Which publishers are participating?

The list of endorsing publishers is growing and includes many if not most of the major commercial and society publishers, http://lockss.stanford.edu/about/titles.htm As of September 2005 120 titles are available for preservation.

Which libraries are participating?

The list of libraries is growing LOCKSS boxes are deployed at almost 100 libraries worldwide, http://www.lockss.org/about/users.htm

What is needed to take the LOCKSS program to the next level?

With the community's help, we have accomplished what some had considered impossible: we have a working digital preservation tool for web-based materials, and a growing community of publishers and librarians. A tremendous amount of work lies ahead, specifically we need to:

- **Grow the community of participating libraries** to ensure there are enough copies of any one piece of content to "keep it safe" and to share the costs of continuing to develop and maintain the software.
- **Build a cooperative collection development community** to encourage publishers' participation. Specifically, we need to use the power of the market with subscription-based publishers and to share the relatively high cost of obtaining IP permission from small, yet important open-access publishers.
- **Support continued development** to allow preservation of a broader range of content, support a broader range of hardware, further reduce the cost of running a LOCKSS box, and further improve the system's reliability and scalability.

What are your plans to sustain the LOCKSS program over the long term?

The program's long-term sustainability requires a small central staff (the Stanford team) for some comparatively short period of time and a growing community-based cooperative effort over the long term. The Stanford team, one librarian and four engineers, will provide support over a transition period and complete this developmental phase while looking for a long-term home. The Stanford team will foster an open source technical community, transferring knowledge, skills and responsibility for the LOCKSS program from Stanford to the international community through the LOCKSS Alliance.

LOCKSS Alliance

The LOCKSS Alliance is a membership organization open to libraries interested in LOCKSS as part of their strategy for building and preserving digital collections of e-journals and other web-based content. It is governed by a Board of Directors and staffed by project team members.

LOCKSS Alliance Board of Directors

Carol Pitts Diedrichs, University of Kentucky	Michael A. Keller, Stanford University
Nancy L. Eaton, Pennsylvania State University	Susan K. Nutter, North Carolina State University
David S. Ferriero, New York Public Library	Ann Okerson, Yale University
Brinley Franklin, University of Connecticut	Carton Rogers, University of Pennsylvania

Membership fees support ongoing technical development of the software, as well as Alliance activities. Member benefits include early access to LOCKSS strategies, documentation, and pre-release software; implementation workshops for collections and technology; and opportunities to engage directly with the LOCKSS Program Director and technical staff.

Through participation in the LOCKSS Alliance, libraries help shape the scope and content of their LOCKSS assets in line with local priorities, thereby adding value to the library. In addition, members of the Alliance community are offered strategic opportunities to help determine long-term priorities and directions for the evolution of the LOCKSS software and program. Through information sharing, collaboration, and networking, the Alliance coordinates community engagement so individual member activities are magnified and leveraged. A vibrant, active, and engaged user community is key to the success of Open-Source efforts like LOCKSS. By joining the Alliance, you will help ensure that LOCKSS remains a vital and useful resource for years to come.

Information about LOCKSS Alliance member services and benefits, and Alliance organization and governance details can be found on the LOCKSS web site, http://www.lockss.org/alliance/alliance.htm To Join the Alliance, complete the Library Invoice form at http://www.lockss.org/documents/for_libraries.pdf

Victoria Reich
Director LOCKSS Program
650.725.1134
www.lockss.org