April 2006

ATG Interviews Victoria Reich

Cris Ferguson

Furman University, cferguson13@murraystate.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Ferguson, Cris (2006) "ATG Interviews Victoria Reich," *Against the Grain*: Vol. 18: Iss. 2, Article 15.

DOI: <https://doi.org/10.7771/2380-176X.4644>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

ATG Interviews Victoria Reich

Director LOCKSS Program, Stanford University Libraries

<vreich@stanford.edu> <http://www.lockss.org/vicky.htm>

by **Cris Ferguson** (Electronic Resources/Serials Librarian, James B. Duke Library, Furman University, 3300 Poinsett Highway, Greenville, SC 29613; Phone: 864-294-2713) <cris.ferguson@furman.edu>

Column Editor's Note: In his column for this issue of *ATG* Greg Tananbaum profiles **LOCKSS** and **Portico**, describing how the programs work and highlighting the differences and similarities between the two. Recently, **Eileen Fenton**, Executive Director of **Portico**, and **Vicky Reich**, Director of the **LOCKSS Program**, very kindly took some time to answer a few questions for me, going into a little more detail about their organizations. In these interviews we touch on a variety of subjects including the sustainability of archiving, quality control, and new developments at the two organizations. **Reich** also took some time to explain **CLOCKSS**, the new **LOCKSS** initiative. — CF

ATG: How long have you been with the **LOCKSS Program** and what is your role?

VR: I helped to co-found the **LOCKSS Program** with A LOT of "help from my friends," most notably **David S.H. Rosenthal**, Chief Scientist **LOCKSS Program**.

When we started the **LOCKSS Program** (www.lockss.org) I was Assistant Director of **HighWire Press**. Two of my many responsibilities were assisting publishers think through issues relating to online/print subscription models and managing a very large study on how people use online electronic journals. These experiences plus years as head of serials and acquisitions (**Stanford University**, **National Agricultural Library**) and as a reference librarian (**Library of Congress**, **University of Michigan**) influenced the design of the **LOCKSS Program**. And most recently I became involved with the **CLOCKSS** initiative, a community managed dark archive.

ATG: Who created the **LOCKSS Program**?

VR: So many people and groups...it was a community effort. The **LOCKSS Program** received several NSF grants and two grants from the **Mellon Foundation**, and considerable support from **Sun Microsystems**, **Stanford University**, **Intel Laboratories**, and the **Hewlett Packard Laboratories**. Six libraries alpha tested the **LOCKSS Program** from 1998 to 2001: **Harvard University**, **Columbia University**, **U. California Berkeley**, **Stanford University**, **University of Tennessee**, and **Los Alamos National Laboratory**. Thirty libraries beta tested the software from 2002 through mid-2004. The community is now running approximately 140 **LOCKSS** boxes.

ATG: What need does the **LOCKSS Program** fill?

VR: The **LOCKSS** system reinstates the traditional role of librarians by allowing libraries to fulfill their responsibility to take custody of and preserve cultural and social assets for future generations.

Ten years ago Web technology forced a change in the business rela-

tionship between librarians and publishers. Libraries could no longer take custody of materials — they now lease subscription materials or "just access" non-subscription materials. It disrupted the role libraries have played in society for hundreds of years as trusted keepers of information and culture for future generation.

The **LOCKSS** system automatically ingests content as part of the subscription process, robustly preserves and migrates the content to new formats, and transparently provides access to local users whenever the material is not available from the publisher's server.

ATG: What has been the reaction of publisher's and librarians since the system was released into production in 2004?

VR: Take a look here <http://lockss.stanford.edu/about/titles.htm>, and here <http://lockss.stanford.edu/about/users.htm>.

Over 70 publishers have chosen to preserve their materials in libraries using the **LOCKSS** system. This uptake was completely through word of mouth and community involvement.

In addition to running **LOCKSS** boxes to preserve electronic journals, libraries and consortia groups are preserving an incredibly WIDE variety of content [image collections, Websites, archival and manuscript collections]. They are working to preserve databases, blogs and books. **GPO** is leading a federal government depository library document preservation project and states are getting into the game as well. Remarkable what a community and a bit of open source software can accomplish.

ATG: How is the **LOCKSS Program** currently funded, and what is your business plan? What do you think the long term sustainability of this model is?

VR: We are moving toward full sustainability and ending reliance on "soft money" via the **LOCKSS Alliance**. In the first year of the **LOCKSS Alliance** the community has provided 2/3 of what's needed for full sustainability. We keep costs low. The **Stanford** team is small, lean, and extremely efficient. By policy, the **LOCKSS** team at **Stanford** will not increase. As the need for technical expertise increases we are growing an open source technical community. A centralized technical staff is a vulnerable point of failure for a wide variety of reasons. The community is our marketing vehicle. The **LOCKSS Program** approach has gained world wide adoption via word of mouth and "neighbor" recommendations. The **LOCKSS** board* and technical policy committee** provide governance. As central costs are expected to remain constant, and the **LOCKSS Alliance** membership is expected to grow —> **LOCKSS Alliance** membership dues will decrease over time.

The **LOCKSS Alliance** Board

- **Carol Pitts Diedrichs**, Dean of Libraries, **University of Kentucky Libraries**
- **Nancy L. Eaton**, Dean of University Libraries, **The Pennsylvania State University**
- **David S. Ferriero**, Director and Chief Executive of the Research Libraries, **New York Public Library**
- **Brinley Franklin**, Vice Provost for University Libraries, **University of Connecticut**
- **Michael A. Keller**, University Librarian, Director of Academic Information Resources, **Stanford University**
- **Susan K. Nutter**, Vice Provost and Director of Libraries, **North Carolina State University**

continued on page 51

ATG Annual Survey Report from page 46

Libraries at 21%; *Publishers Weekly* at 16%; *Special Libraries* at 15%; *LACTS* at 6%; *Wired* at 8%; *Internet World* at 4%; and 56% selected "other."

Among the other journals listed, *American Libraries*, *Journal of Academic Libraries*, *Journal of American Medical Libraries* and *LRTS* each came in with several votes.

Thank you to all who participated. Your input is greatly appreciated! 🐘

- Ann Okerson, Associate University librarian, Yale University
- Carlton Rogers, Vice Provost and Director of Libraries, University of Pennsylvania
- Chris Rusbridge, Director, Digital Curation Centre, U.K.

**The LOCKSS Alliance
Technical Policy Committee**

- Michael Seadle, Michigan State University
- Beth Sandore, University of Illinois at Urbana-Champaign
- Geneva Henry, Rice University
- Martin Halbert, Emory University
- Nancy McGovern, Cornell University

ATG: Could you tell ATG readers a little bit about the **CLOCKSS** project?

VR: A group of publishers, librarians, and learned societies launched an initiative employing the **LOCKSS** technology to support a community-managed “large dark archive” that serves as a failsafe repository for scholarly content. **Controlled LOCKSS (CLOCKSS)** aims to provide the global research and scholarly community perpetual access to journal content, for orphaned or abandoned content and in the event of a long-term business interruption.

Charles Henry, Vice Provost and University Librarian at **Rice University**, has said “**CLOCKSS** is a critical initiative for librarians. It is managed by the community and all members share the common goal of sustaining the scholarly record. **CLOCKSS** continues the stewardship that research libraries have collectively played for the printed format; continuing this role is a keystone in society’s transition to digital materials.”

ATG: How many libraries and publishers are participating in **CLOCKSS**, and what is their role in the project?

VR: The **CLOCKSS** initiative is a community-managed membership organization of libraries and publishers. Libraries and publishers govern the **CLOCKSS** initiative as equal partners. One of the strengths of the **CLOCKSS** initiative is that all participating organizations have a long history of survival and members understand issues of long-term sustainability.

Libraries and Date Founded:

Edinburgh University — 1582
Indiana University — 1820
New York Public Library — 1895
Rice University — 1912
Stanford University — 1891
University of Virginia — 1825

Publishers and Date Founded:

American Medical Association — 1847
American Physiological Society — 1887
Blackwell — 1897
Nature Publishing Group — 1869
Oxford University Press — 1478
SAGE Publications — 1965
Springer — 1842
Taylor and Francis — 1798
John Wiley & Sons — 1807

Elsevier (which is participating in all discussions and is sharing in financial support) — 1880

The **CLOCKSS** member libraries and publishers are sharing

continued on page 52

**CROSSREF
DRIVES
ONLINE
RESEARCH.**

Publications around the world
rely on the reference linking
power of CrossRef.

CrossRef links millions
of articles and chapters,
from thousands of
publications, from
hundreds of publishers
and societies.

So reference linking works.
Permanently.

Take a tour at
www.crossref.org.

crossref.org

DOIS FOR RESEARCH CONTENT

CLOCKSS initiative expenses equally which includes money for additional servers, support staff and development costs.

ATG: *What are the differences between the LOCKSS Program and the CLOCKSS Initiative?*

VR: The main difference between the **LOCKSS Program** and the **CLOCKSS** initiative is that **LOCKSS** provides a community approach to long term preservation of a library's local collections while **CLOCKSS** aims to provide a long-term global archiving solution that will serve the joint library and publisher communities in the event of a long-term business interruption or in making orphaned or abandoned works readily available to the scholarly community.

In **LOCKSS**, librarians use their **LOCKSS** boxes to collect and preserve the journal content locally that they subscribe to. With the publisher's permission **LOCKSS Alliance** libraries no longer "just lease content." Publishers have control over which libraries take custody of what materials and when this occurs. Preserved materials are available to the local community when the publisher is not able to resolve a specific URL request.

In **CLOCKSS**, libraries preserve content from member publishers to which they subscribe and content they don't. **CLOCKSS** content would only be available after a "trigger" event, such as the material no longer being available from the publisher. In these situations, the publishers, librarians, and representing societies begin the collaborative process to determine whether materials should be made generally available to all for a limited or an indefinite period of time.

LOCKSS has a large number of participating libraries; it allows a library to locally preserve its own subscription. **CLOCKSS** has a limited number of library participants; the dark archives will be held on behalf of the broader community.

ATG: *What will be done with the results and findings with the results of the CLOCKSS initiative?*

VR: The findings of the **CLOCKSS** initiative will be shared with the community as we move forward for comment and feedback. The result of the **CLOCKSS** initiative will be a robust community managed archive, open to all publishers.

ATG: *How does the LOCKSS technology ensure timely and accurate receipt of publisher data?*

VR: The **LOCKSS** system ensures "timely receipt" of data via an "html publisher manifest page." The publishers put online a manifest page volume by volume. The content is collected as the title is published. The **LOCKSS** technology works in the same way as other systems, the publisher's cooperation is required. If a publisher refuses or neglects to put online a manifest page — no new content is preserved. This is the same as a publisher refusing or neglecting to send a paper journal to a library or a file of data to a centralized archiving service. In the **LOCKSS** system however, all libraries that are preserving a title see if a publisher has "dropped out" and can rally to apply market pressure. The larger publishers have automated the publisher manifest page process.

ATG: *How does the LOCKSS technology ensure data accuracy? What are the control mechanisms?*

VR: The data must be confirmed to be the same as the publisher published. Each **LOCKSS** box collects content independently from the publisher's Web server. Each **LOCKSS** box then compares the content it has collected with other **LOCKSS** boxes and with the publisher's Website and algorithmically determines an authoritative version. The authoritative version is central to the continual audit and repair process. Digital information is fragile: "1"s continually change to "0"s. These changes are not detectable "by eye" — until it's too late and the file is corrupted. The **LOCKSS** protocol (the preservation layer) performs this continual audit and repair. The research underpinning the **LOCKSS** protocol won a prestigious **ACM** research award in 2004. Digital information must be migrated to new formats when brows-

ers readers are using can no longer understand that format. The **LOCKSS** system converts Web content from one format to a newer format with a process called "migration on the fly." Migration occurs only when content is requested and is transparent to the reader. <http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html>

"Format migration on the fly" enables the **LOCKSS** system to:

- preserve the original look-and-feel, which in the case of increasing amounts of content is a large part of the value
- reduce the cost of ingest, allowing more material to be preserved per dollar.
- postpone the inescapable costs of migration, taking advantage both of the time value of money, and of the technology cost curve.
- only migrate material when the reader requires it, vastly lowering the amount of content that needs to be processed
- allow what the reader sees to be the result of the best available format migration technology at the time access is required.

ATG: *What do you think are some of the biggest obstacles to the permanent archiving of journal content, and what type of solution do you think will come the closest to surmounting those obstacles?*

VR: The single greatest threat to materials being preserved over the long term is money. Societies will have good times and bad. Keeping content safe must be a marginal expense in order to decrease the threats during bad times as well as to maximize available funds for new acquisitions during good times.

The **LOCKSS** system does this by minimizing processing and infrastructure costs

- a. Don't touch the data on ingest. The publishers have spent a lot of money processing the data. It's "cooked" by the time it's published.
- b. Migrate data when needed for access ("on the fly")
- c. Leverage the Web as the access delivery platform, use infrastructure that society as a whole is maintaining.

The second biggest threat to digital preservation is technical arrogance. Open source software is critical base of a long-term digital preservation system. The **LOCKSS** community is building and using open source software. Many eyes and minds are examining and contributing to the software; many eyes are confirming processing claims and helping to correct INEVITABLE SOFTWARE BUGS. The system is fully documented and the entire source code is available from sourceforge.net. No technical team is infallible, and limited scheduled software audits are not sufficient to overcome this weakness. One-time auditors can never know the system in sufficient depth.

The third biggest threat to digital preservation is "insider attacks." Industry experience shows most security breaches are from people with authorized access. In a centrally ADMINISTERED archiving solution a change to the "mother file" will just be propagated to the back up files — and no one will know. Content with potential economic value (patents, FDA approval) or with real or perceived political volatility (stem cell research) is particularly vulnerable. The **LOCKSS** boxes are independently administered repositories. Authorized administrators of one **LOCKSS** box have no access to others elsewhere.

ATG: *Where do you see the LOCKSS Program and the CLOCKSS Initiative heading in the next year? Five years?*

VR: Both are powerful community managed approaches towards solving an important societal issue, the preservation of today's materials for tomorrow's world citizens. In general, it's bad for society to concentrate tools and resources in the hands of a few powerful institutions. We prevent this by providing transparency in process, transparency in legal documentation, transparency in finances, and open source software for the community to examine and use.

As **Gordon Tibbitts** President of **Blackwell Publishing** has said, "A solution built by the community will gain the broadest level of support and trust."