



<http://www.lockss.org>

L O T S O F C O P I E S K E E P S T U F F S A F E

*M. Elizabeth Cowell, Stanford University Libraries
Depository Library Council Meeting, Denver, CO*

April 15, 2007





A Library

Serves local and global communities

- Collects documents
- Provides access
- Preserves artifacts for future use

is a MEMORY organization





Print Libraries

Resist attacks

- Natural disasters
- Ideology
- Government swings in policy and funding

Self healing

- Inter library loan
- Preservation copies

Are a preservation system





Web's Impact on Libraries

Libraries now

- Lease subscription materials
- Access free materials

Libraries unable to

- Own collections
- Fill memory role





What is a library with no collections?

How do we preserve open access literature and information in the public domain?

How do we guard against losing access when subscriptions are cancelled or titles ceased?





LOCKSS

- A persistent access preservation system for web delivered information
- Uses open source software
- Based on the print library model





LOCKSS

Libraries acquire copies of important content

- Keep copies on shelves
- Give access to local readers and supply to other libraries:
 - readers can easily find a copy
 - a “bad guy” has trouble finding and destroying all copies





LOCKSS Boxes

- Crawl and collect HTTP content
 - All formats (PDF, HTML, JPEG, TIF, Audio, Video)
- Preserve content integrity
 - Independent collection
 - Cooperate to audit and repair damage
- Provide access
 - Via web browser
 - Content is never “dark”





Look and Feel to Readers

Example

- PNAS table of contents page
 - from web (9/11/02)
 - from LOCKSS Box

Look and feel is as close as possible
to what the publisher published
Preserved content & “presentation”





Want to be notified by email when new content goes on-line? [\[Sign up for eTOCs\]](#)

Other Issues: [←](#) [→](#)

Table of Contents: Jan 2 2001; 98 (1) [\[Index by Author\]](#) [\[Cover\]](#)

↓ [COMMENTARIES](#)

↓ [PERSPECTIVES](#)

↓ [Physical Sciences:](#)

[Mathematics](#)
[Statistics](#)

↓ [Social Sciences:](#)

[Anthropology](#)

↓ [Biological Sciences:](#)

[Biochemistry](#)
[Biophysics](#)
[Cell Biology](#)
[Developmental Biology](#)
[Ecology](#)
[Evolution](#)
[Genetics](#)

[Immunology](#)
[Medical Sciences](#)
[Microbiology](#)
[Neurobiology](#)
[Physiology](#)
[Plant Biology](#)

↓ [CORRECTIONS](#)

Find articles in this issue containing these words:

[\[Browse & Search All Issues\]](#)

To see an article, click its [Full Text] link. To review many abstracts, check the boxes to the left of the titles you want, and click the 'Get All Checked Abstract(s)' button. To see one abstract at a time, click its [Abstract] link.



PNAS Online

[HOME](#) [HELP](#) [FEEDBACK](#) [SUBSCRIPTIONS](#) [ARCHIVE](#) [SEARCH](#) [TABLE OF CONTENTS](#)

Want to be notified by email when new content goes on-line? [\[Sign up for eTOCs\]](#)

Other Issues: [←](#) [→](#)

Table of Contents: Jan 2 2001; 98 (1) [\[Index by Author\]](#) [\[Cover\]](#)

↓ [COMMENTARIES](#)

↓ [PERSPECTIVES](#)

↓ [Physical Sciences:](#)

[Mathematics](#)
[Statistics](#)

↓ [Social Sciences:](#)

[Anthropology](#)

↓ [Biological Sciences:](#)

[Biochemistry](#)
[Biophysics](#)
[Cell Biology](#)
[Developmental Biology](#)
[Ecology](#)
[Evolution](#)
[Genetics](#)

[Immunology](#)
[Medical Sciences](#)
[Microbiology](#)
[Neurobiology](#)
[Physiology](#)
[Plant Biology](#)

↓ [CORRECTIONS](#)

Find articles in this issue containing these words:

Enter

[\[Browse & Search All Issues\]](#)



Benefits

- Spreads responsibility for content among various locations and administrations
- Less risk of catastrophic failure
- Different communities can have their own collections based on their needs
- Provides public access





What is LOCKSS up to?

More than just e-journals

- Alabama Academic Libraries grant
- Alaska State Documents
- Fugitive Documents Network
- CLOCKSS
- OpenLOCKSS Project
- UK LOCKSS Pilot Programme





GPO LOCKSS Pilot

http://www.access.gpo.gov/su_docs/fdlp/lockss/index.html





Pilot Partners

- Alaska State Library
- Arizona State University
- Brigham Young University
- Columbia University
- Dartmouth College
- Deutsche Bibliothek
- Georgetown University
- Georgia Tech
- Indiana University
- National Agricultural Library





Pilot Partners

- North Carolina State University
- Portland State University
- Rice University
- Stanford University
- University of Connecticut
- University of Kentucky Libraries
- University of Tennessee
- University of Utah
- University of Wisconsin-Madison
- U.S. Government Printing Office
- Yale University Law Library





Pilot Final Report

What was the pilot intended to test?

Distribution

What was the pilot NOT intended to test?

Preservation





Problems highlighted in report

Recall

Alliance membership

GPO staff time





Value of the Pilot from the partners' perspective

- Tested the distributed model of digital library collections
- Affirmed Depository Libraries' role in ensuring permanent public access
- Showed how easy and efficient LOCKSS is to use





Possible roles in a LOCKSS-powered e-FDLP

Libraries

- Select, organize, enhance access, provide service, preserve

GPO

- Identify, capture, describe, authenticate, affirmatively disseminate Federal web-based content





Partners in Preservation

- Depository libraries must demand to be part of the preservation system of the FDLDP
- Say **YES** on the Biennial Survey
- The risk of loss when all government information on government servers is too high



Future for LOCKSS-DOCS?

LOCKSS is continuing to discuss government information





Would you like to join the
conversation?

Cowell@stanford.edu



LOCKSS at Georgia Tech

LOCKSS-based initiatives included in current 5-year Strategic Plan; will also be noted in new Strategic Plan.

The Library joined the LOCKSS Alliance in spring 2004, noting that it is a “major step forward in the Library’s consortial approach to bring Georgia Tech faculty, staff, and students scholarly digital content, sustain its availability over time, and achieve cost savings”.

GPO LOCKSS Pilot Project announcement (May 2005) sent to LOCKSS Alliance members. The Library had been involved in the MetaArchive project (a partnership between LC and 6 ASERL members) that is a private LOCKSS network.

For LOCKSS, used an old PC (4-5 years old); purchased a larger hard drive for about \$200.

As GPO’s Final Analysis of the GPO LOCKSS Pilot Project notes (p. 7), “pilot partner LOCKSS boxes automatically downloaded the plug-in from Stanford. This action also initiated the initial harvest and subsequent crawls of the e-journal issues”. *The Library’s Systems Dept. staff time was not needed.*

The Library is willing to share plug-in expertise with FDLP libraries.

LOCKSS has become a core technology for the Library; it is our primary strategy for digital preservation.

SUGGESTIONS

1. Use the LOCKSS software – it’s open source – your library does NOT have to be an Alliance member (the members help support R&D, etc.)
2. Consider learning opportunities for using LOCKSS technology
- example: Distributed Digital Preservation Workshop, May 30-June 1, 2007 – at Emory University in Atlanta (\$150 for one day; \$350 for 3 days).
Information available at: www.metaarchive.org/
3. Encourage your library’s administration to include using LOCKSS technology for preservation of your library’s collections.
4. Explore consortial opportunities to use LOCKSS --
- cooperative agreements at the local, state, or regional levels

Patricia Kenly,
Georgia Institute of Technology
April 2007

LOCKSS at Georgia Tech

Patricia E. Kenly
April 2007

LOCKSS Alliance

- The Library joined the Alliance in spring 2004:
 - “a major step forward in the Library’s consortial approach to bring Georgia Tech faculty, staff, and students digital content, sustain its availability over time, and achieve cost savings”.

The Library’s 1995-2005 Effectiveness Report

Strategic Plan

- LOCKSS-based initiatives included in current strategic plan
- Will also be noted in new strategic plan; a primary part of the Library's agenda in building a digital repository for the collections

INITIATIVES

- MetaArchive Project (www.metaarchive.org/)
 - Partnership between:
 - Auburn University
 - Emory University
 - Florida State University
 - Georgia Tech
 - Library of Congress
 - University of Louisville
 - Virginia Tech

Part of NDIIPP (National Digital Information
Infrastructure and Preservation Program)

INITIATIVES

- MetaArchive project - functions as a private LOCKSS network
- Library operates a general LOCKSS server for e-journals
- GPO LOCKSS Pilot Project

GPO LOCKSS PILOT PROJECT

- Announcement sent to LOCKSS Alliance members (May 2005)
- Georgia Tech had positive experience with LOCKSS
 - interested in applications to government documents journals
- Used an old PC; purchased a larger hard drive (about \$200)

GPO LOCKSS PILOT PROJECT

- Minimal staff time
 - Evaluated journals for consideration in project
 - Monthly conference calls with partner institutions
 - Real World Scenario – librarian met briefly with Systems staff

TECHNOLOGICAL ISSUES

- Library's Systems Dept. employs a grad student for LOCKSS-based initiatives
- In response to the concern raised in the Final Analysis re technological capabilities (p. 11), Georgia Tech is willing to share expertise with FDLP libraries

LOCKSS - OUTCOMES

- LOCKSS has become a core technology for the Library
- LOCKSS is the Library's primary strategy for digital preservation

PLANS OF ACTION

- Use the LOCKSS software –
 - it's open source
- Consider learning opportunities for using LOCKSS technology
 - Upcoming workshop at Emory Univ. (May 30-June 1) – information available at *www.metaarchive.org/*

PLANS OF ACTION

- Encourage administration at your library to include using LOCKSS technology for preservation of your library's collections
 - Consider adding this to library's Strategic Plan
- Explore consortial opportunities to use LOCKSS
 - Cooperative agreements at the local, state, or regional levels

Patricia E. Kenly
patricia.kenly@library.gatech.edu
404-894-1389

Preserving Digital Information: LOCKSS

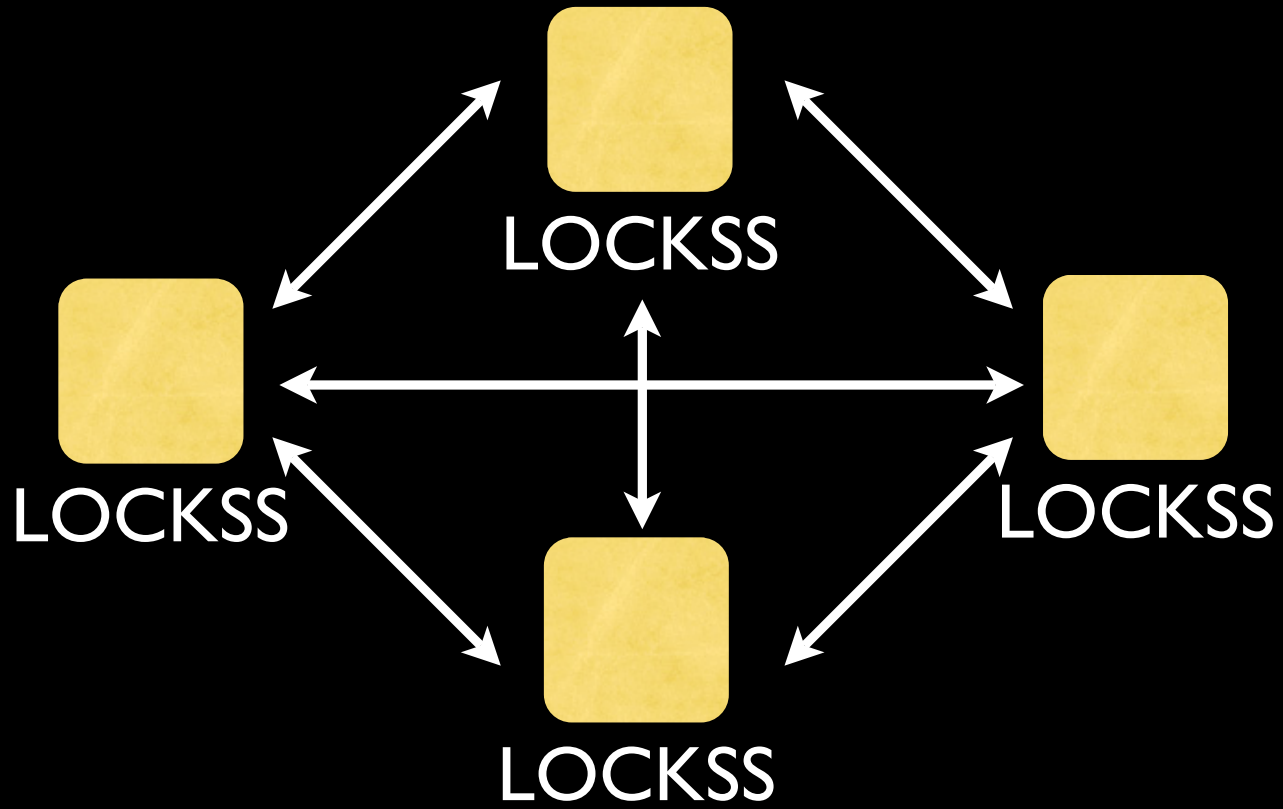


What is LOCKSS?

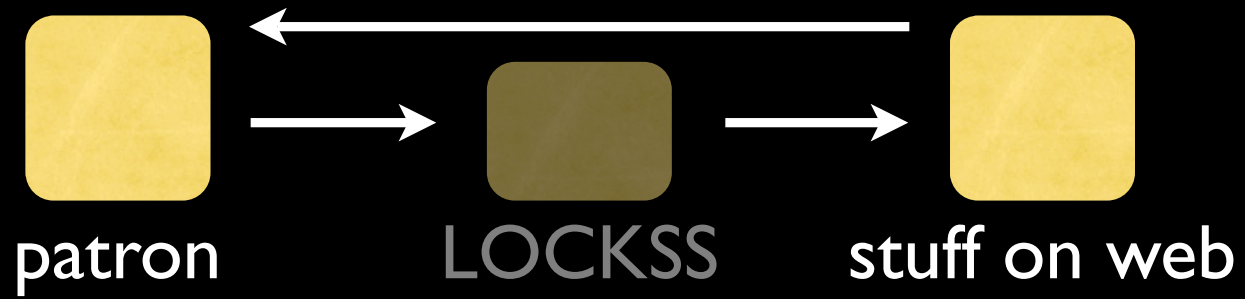
Lots
Of
Copies
Keeps
Stuff
Safe

Three pictures
to keep in mind.

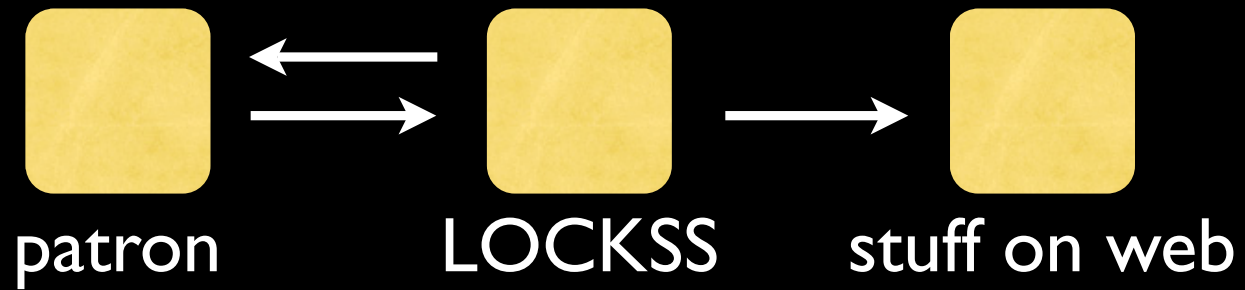
One



Two



Three



Why do we need it?
(or something like it)

How it works

Three basic functions

- harvesting
- preserving
- displaying

Harvesting

- a publisher creates a manifest page and a plug-in
- the plug-in is distributed through the LOCKSS network
- a local cache administrator (that's you) checks a box next to the plug-in
- the cache crawls the publisher's site according to the directions in the plug-in

Pictures.

How LOCKSS gets stuff



LOCKSS gets instructions (plug-in) from network.

How LOCKSS gets stuff



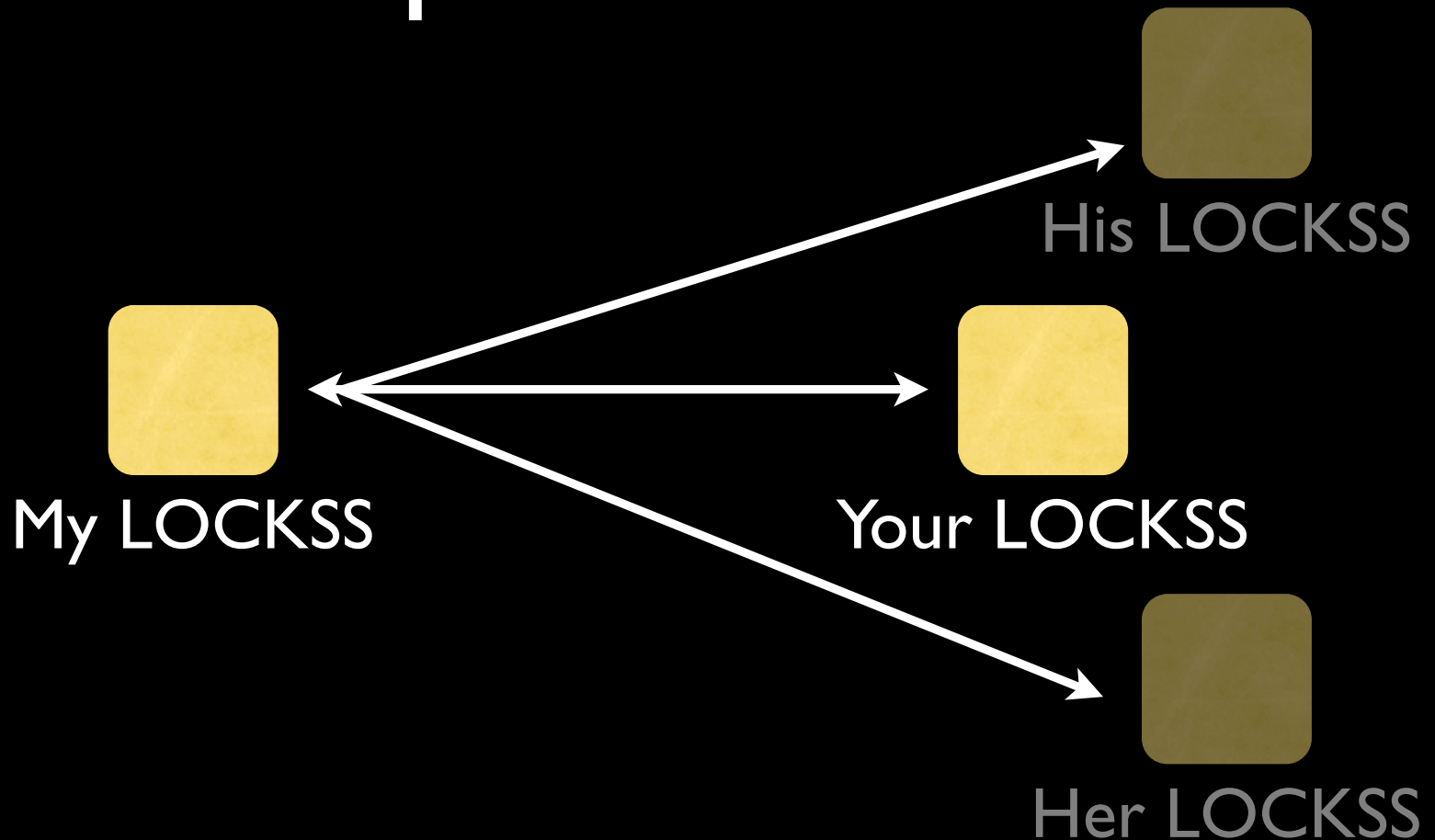
LOCKSS cache follows plug-in instructions
on what to get and where to find it.

Preserving

- a cache polls its peers to determine if its copies match the other caches' copies
- when enough caches reply and vote on the correct hash, each knows that it has a good copy
 - or not
 - the losing cache repairs its files

Picture.

How LOCKSS keeps stuff safe



Displaying

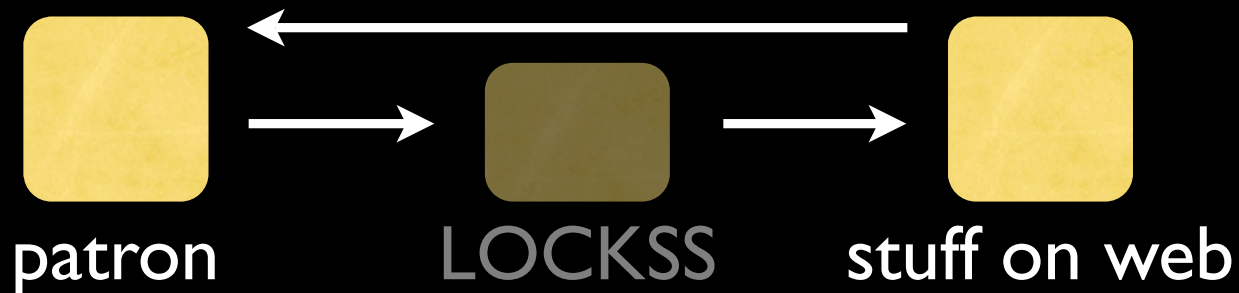
- each local cache is also a proxy server
- when properly configured, LOCKSS is transparent to the patron
- when a patron asks for a title that is preserved in your cache, the box intercepts the request and checks to see if the content is still available from the publisher
 - if yes, then patron gets content from the publisher's servers
 - if no, then LOCKSS cache seamlessly provides the content from its local copy

Pictures.

How patrons get stuff

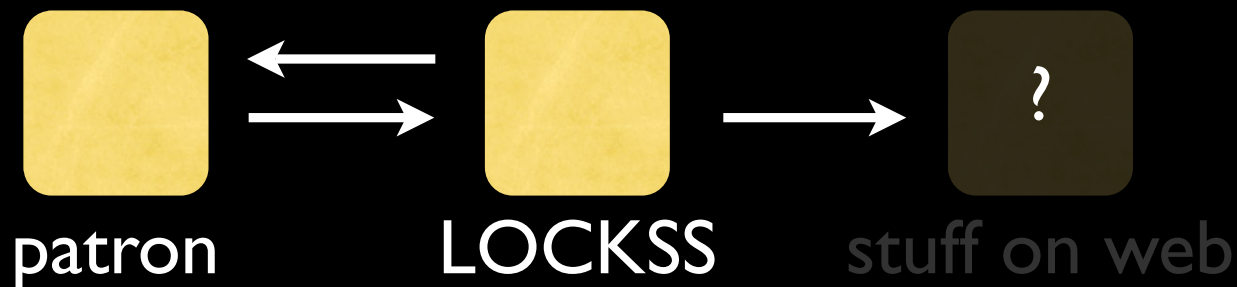


How patrons get stuff with LOCKSS



Stuff still on the publisher's web site.

How patrons get missing stuff with LOCKSS



Stuff gone from web site or web site gone.

Three thoughts
to take away.

If you are a librarian,
you only need
to worry about
a one-time setup.

LOCKSS is industrial
strength computer science,
but you don't have
to be a geek to use it.

cheap resources
+
library cooperation

a robust system

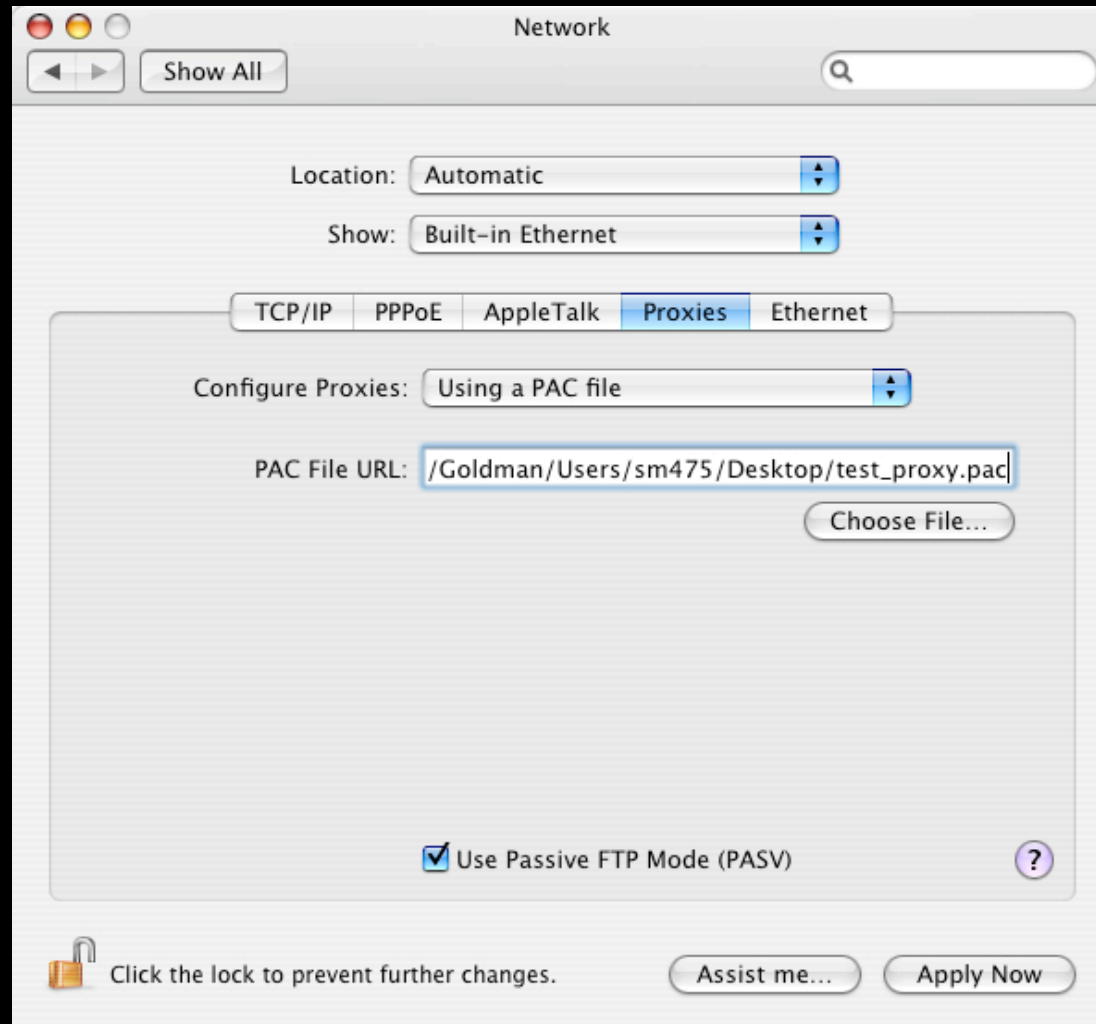
What it looks like.



Safari can't connect to the server.

Safari can't open the page
"http://permanent.access.gpo.gov/lockss/NIST_Journal_of_Research/volume110/index.html"
because it could not connect to the server "permanent.access.gpo.gov".





This copy was downloaded from the agency's web site by the U.S. Government Printing Office under Title 44 USC, 07/05/2005. External links, forms, and search boxes may not function within this collection.

Journal of Research of the **National Institute of Standards and Technology**

Volume 110

Number 1

January-February 2005

[Front Cover—Title Page—Contents](#)  1497 kb

Articles

[Details of the 1998 Watt Balance Experiment Determining the Planck Constant](#)  2324 kb

Richard Steiner, David Newell, and Edwin Williams

[Re-Analysis of the Uncertainty of the 0.895 \$\mu\text{m}\$ Diameter \(NIST SRM® 1690\) and the 0.269 \$\mu\text{m}\$ Diameter \(NIST SRM® 1691\) Sphere Standards](#)  523 kb

G. W. Mullholland

[Slip Correction Measurements of Certified PSL Nanoparticles Using a Nanometer Differential Mobility Analyzer \(Nano-DMA\) for Knudsen Number From 0.5 to 83](#)  1407 kb

Jung Hyeun Kim, George W. Mulholland, Scott R. Kukuck, and David Y. H. Pui

[Measurement of Workability of Fresh Concrete Using a Mixing Truck](#)  1076 kb

Sofiane Amziane, Chiara F. Ferraris, and Eric P. Koehler

Cache administration.



™

Cache Administration

at 18:03:58 08/17/06, up 8d5h40m

[Journal Configuration](#)
[Admin Access Control](#)
[Proxy Access Control](#)
[Proxy Options](#)
[Proxy Info](#)
[Daemon Status](#)
[Contact Us](#)
[Help](#)

Welcome to the administration page for LOCKSS cache **read2.law.yale.edu**.

Journal Configuration	Add or remove titles from this cache
Manual Journal Configuration	Manually edit single AU configuration
Admin Access Control	Control access to the administrative UI
Proxy Access Control	Control access to the preserved content
Proxy Options	Configure the audit proxy and the ICP server.
Proxy Info	Info for configuring browsers and proxies to access preserved content on this cache
Daemon Status	Status of cache contents and operation
Help	Online help, FAQs, credits

L O T S O F C O P I E S K E E P S T U F F S A F E™

Daemon 1.18.3 built 02-Aug-06 19:40:46 on build3.lockss.org, OpenBSD CD 182



Journal Configuration

[Journal Configuration](#)

[Manual Journal](#)

[Configuration](#)

[Admin Access Control](#)

[Proxy Access Control](#)

[Proxy Info](#)

[Daemon Status](#)

[Hash CUS](#)

[Logs](#)

[Thread Dump](#)

[Contact Us](#)

[Help](#)

Select one or more collections of titles to add, then click Select Titles.

All Titles (66)

All BioOne Titles (56)

All Free Titles (10)

All Project Muse Titles (0)

LOCKSS Card (0)

Select Titles

[Back to Journal Configuration](#)

LOTS OF COPIES KEEP STUFF SAFE™

Daemon 1.8.2 built 02-May-05 09:23:32 on build3.lockss.org, CD 167



Journal Configuration

- [Journal Configuration](#)
- [Manual Journal Configuration](#)
- [Admin Access Control](#)
- [Proxy Access Control](#)
- [Proxy Info](#)
- [Daemon Status](#)
- [Hash CUS](#)
- [Logs](#)
- [Thread Dump](#)
- [Contact Us](#)
- [Help](#)

Select the AUs you wish to add. Then click Add Selected AUs.

Add Selected AUs

Select All

Clear All

Add?	Archival Unit	Est. size (MB)
<input type="checkbox"/>	Castanea Volume 69	
<input type="checkbox"/>	Cell Stress & Chaperones Volume 9	
<input type="checkbox"/>	The Coleopterists Bulletin Volume 58	
<input type="checkbox"/>	Comparative Parasitology Volume 72	
<input type="checkbox"/>	The Condor Volume 107	
<input type="checkbox"/>	Copeia Volume 4	
<input type="checkbox"/>	Economic Botany Volume 58	
<input type="checkbox"/>	Environmental Entomology Volume 33	
<input type="checkbox"/>	Evolution Volume 59	
<input type="checkbox"/>	Florida Entomologist Volume 87	
<input type="checkbox"/>	Herpetologica Volume 61	
<input type="checkbox"/>	Herpetological Monographs Volume 18	
<input type="checkbox"/>	In Vitro Cellular & Developmental Biology - Animal Volume 40	
<input type="checkbox"/>	In Vitro Cellular and Developmental Biology - Plant Volume 40	
<input type="checkbox"/>	Integrative and Comparative Biology Volume 44	
<input type="checkbox"/>	Journal of Arachnology Volume 32	
<input type="checkbox"/>	Journal of Avian Medicine and Surgery Volume 19	
<input type="checkbox"/>	Journal of Biotechnology Volume 14	

Content Preserved



Daemon Status

at 17:53:38 03/21/07, up 22h15m17s

- [Journal Configuration](#)
- [Admin Access Control](#)
- [Proxy Access Control](#)
- [Proxy Options](#)
- [Proxy Info](#)
- [Daemon Status](#)
- [Hash CUS](#)
- [Logs](#)
- [Thread Dump](#)
- [Contact Us](#)
- [Help](#)

Archival Units

Archival Units

1055 Archival Units
11 awaiting recalc

Volume	Content Size ¹	Disk Usage (MB) ¹	Peers	Polls	Status ²	Last Poll	Last Crawl	Last TreeWalk
Academic Psychiatry Volume 30	93,400,185	103.9	peers	0	Waiting for Poll	00:14:49 10/22/06	21:44:43 03/18/07	never 03/21/07
Accounting History Volume 5	3,125,086	5.4	peers	0	100.00% Agreement	02:30:36 02/12/07	03:07:21 02/27/07	never
Acta Sociologica Volume 48	7,992,857	11.4	peers	0	86.93% Agreement	00:05:26 02/14/07	04:58:15 02/27/07	never
Action Research Volume 3	6,443,633	9.6	peers	0	95.77% Agreement	12:47:23 02/03/07	00:57:54 02/26/07	never
Active Learning in Higher Education Volume 6	5,146,272	7.7	peers	0	Waiting for Poll	12:03:57 10/28/06	21:20:34 03/17/07	never
Adaptive Behavior Volume 13	13,264,401	23.4	peers	0	99.39% Agreement	17:11:49 12/23/06	04:11:53 03/20/07	never