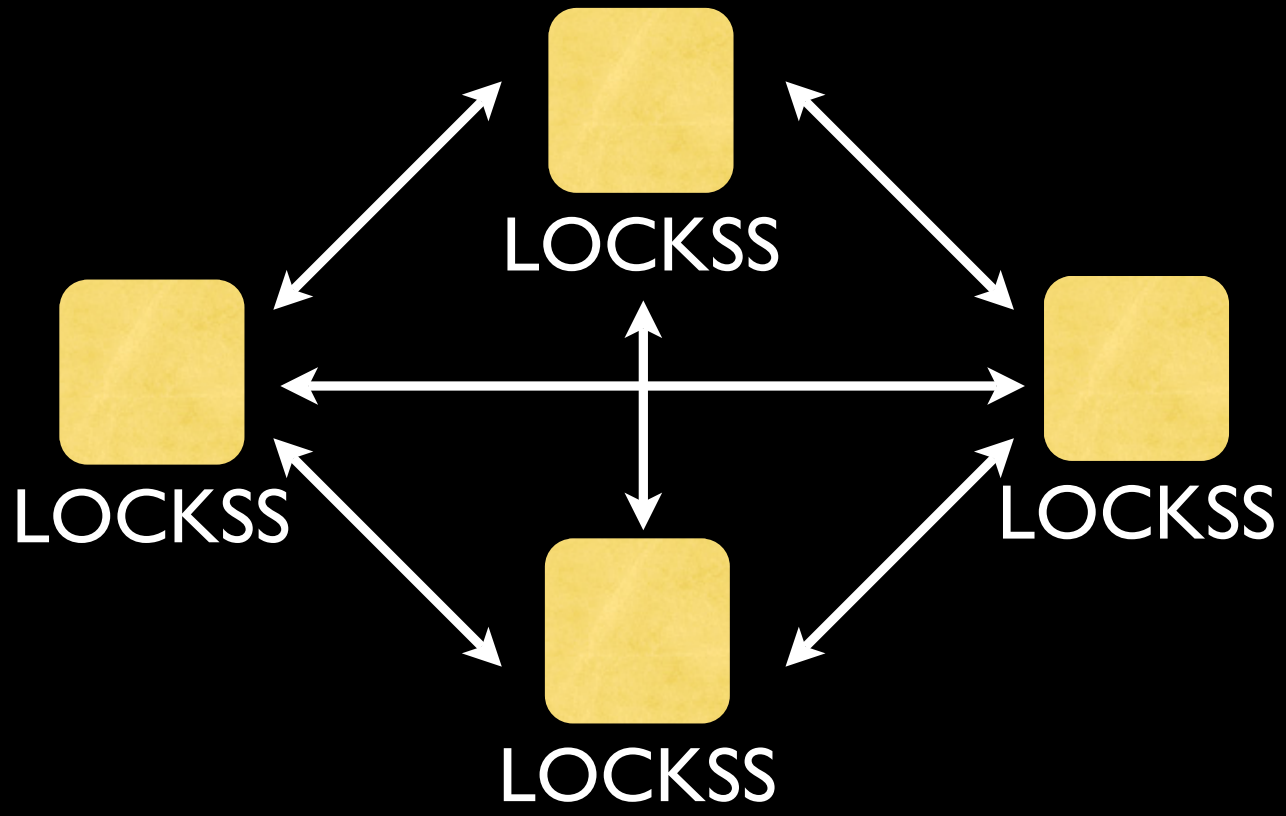
Preserving Digital Information: LOCKSS

What is LOCKSS?

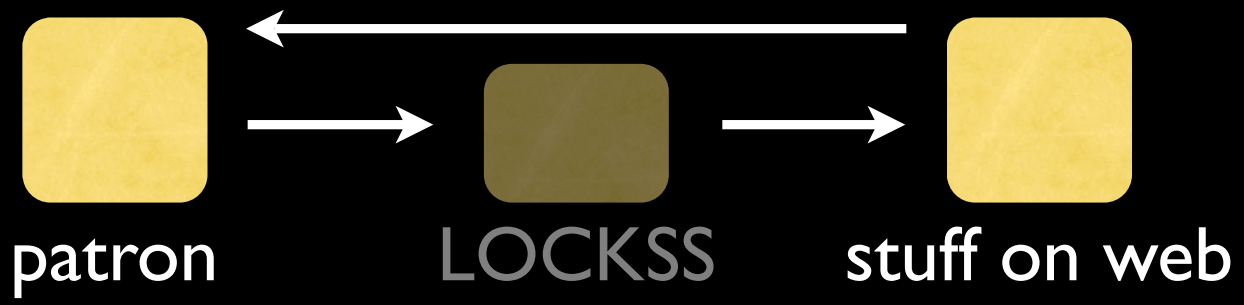
Lots
Of
Copies
Keeps
Stuff
Safe

Three pictures
to keep in mind.

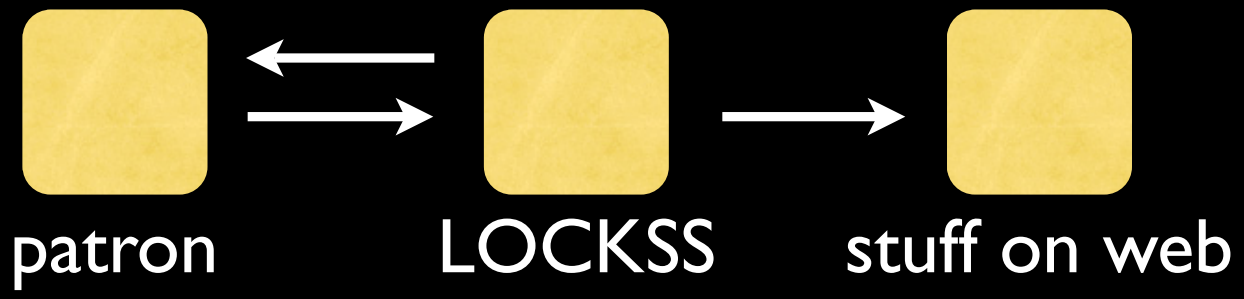
One



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

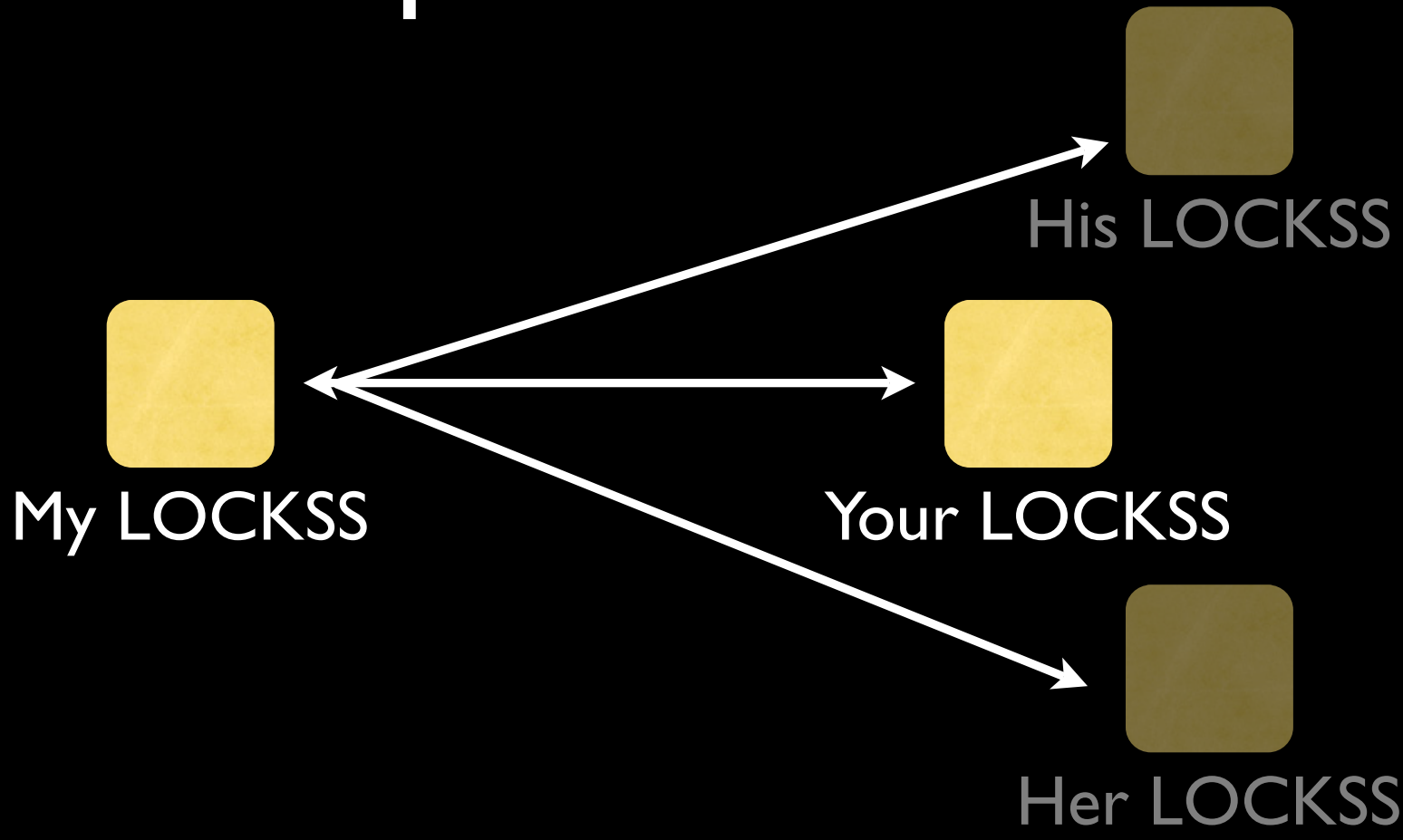
LOCKSS cache follows plug-in instructions on what to get and where to find it.

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

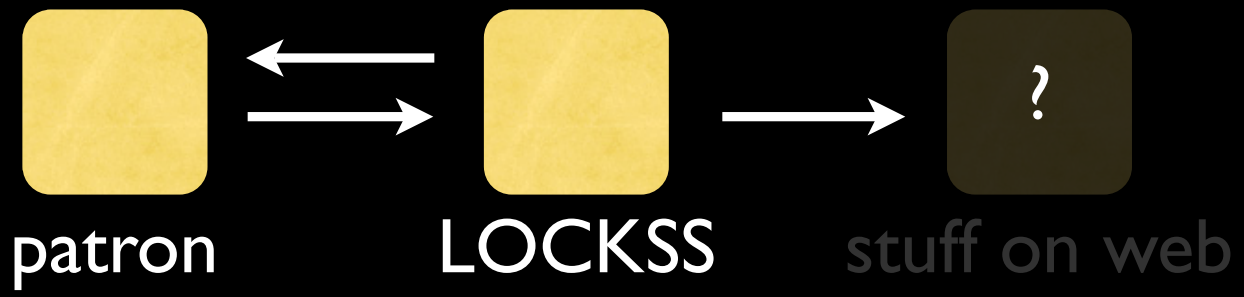
Pictures.

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

+

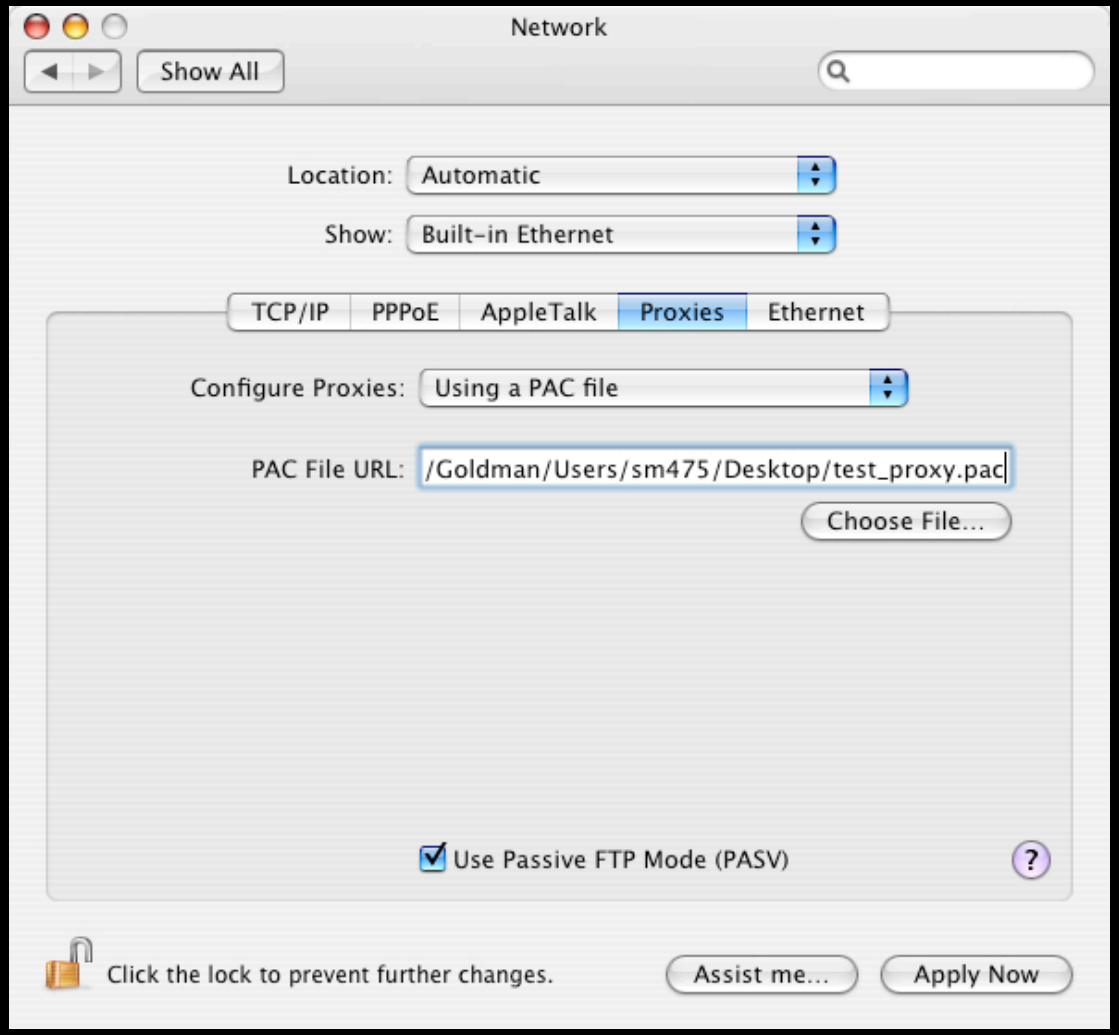
library cooperation

---

a robust system

What it looks like.

Failed to open page
http://permanent.access.gpo.gov/lockss/NIST_Journal_of_Research/volume110/inde
Google
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".


Network
Show All
Location: Automatic
Show: Built-in Ethernet
TCP/IP
PPPoE
AppleTalk
Proxies
Ethernet
Configure Proxies: Using a PAC file
PAC File URL: /Goldman/Users/sm475/Desktop/test_proxy.pac
Choose File...
Use Passive FTP Mode (PASV)
Click the lock to prevent further changes.
Assist me...
Apply Now

JRES 110-1

http://permanent.access.gpo.gov/lockss/NIST_Journal_of_Research/volu

Google

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

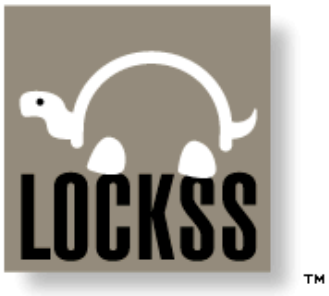
Details of the 1998 Watt Balance Experiment Determining the Planck Constant 2324 kb
Richard Steiner, David Newell, and Edwin Williams

Re-Analysis of the Uncertainty of the 0.895 μm Diameter (NIST SRM® 1690) and the 0.269 μm 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

[redacted] 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

LOTS OF COPIES KEEP STUFF SAFE™

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™

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.

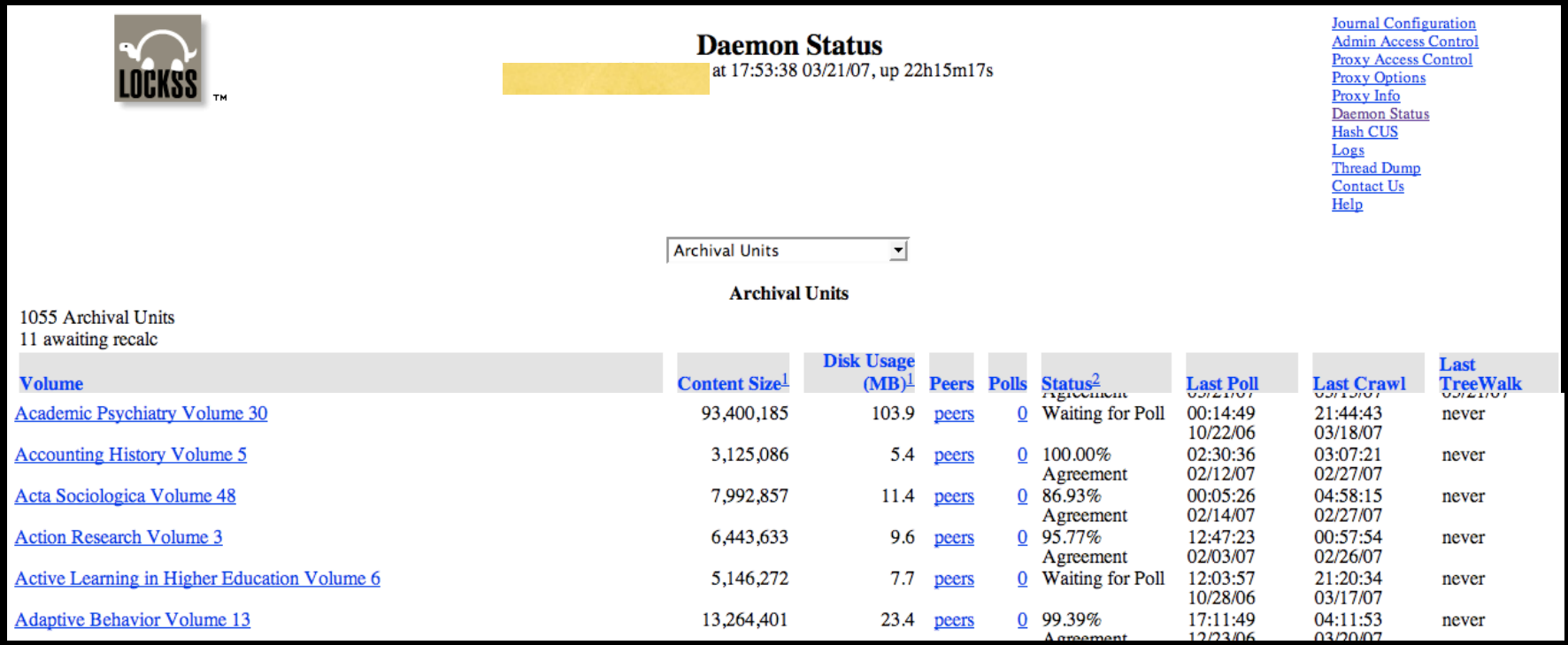
Add Selected AUs

Select All

Clear All

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

Content Preserved



LOCKSS™

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
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