

Preservation Tools For BLOGS

Victoria Reich
Director LOCKSS Program
Stanford University Libraries



www.lockss.org

Introduction

- This talk is in 3 parts
 - About LOCKSS
 - What is it?
 - How does it work?
 - Decentralized preservation
 - Authentication
 - Threats to digital content
 - How blogs could make preservation easier



www.lockss.org

Introduction

- This talk is in 3 parts
 - About LOCKSS
 - What is it?
 - How does it work?
 - Decentralized preservation
 - Authentication
 - Threats to digital content
 - Making preservation easier



www.lockss.org

About LOCKSS

- An international community initiative
- Libraries taking charge
 - To continue their role as information stewards
 - As memory organizations in the digital environment



www.lockss.org

A Bit More Detail...

- Distributed Digital Preservation System
 - Open source peer to peer software
 - Standards: OAIS, OpenURL, HTTP, WARC
- Stanford Libraries - administrative home
- Founded in 1998
- Supported by libraries - LOCKSS Alliance



www.lockss.org

Why Should You Care?

- Librarians use LOCKSS as digital stacks
 - Build and preserve e-collections
 - Subscription content - buy it once and keep it!
 - Open access - not locked up behind fee wall
- Provide 100% perpetual access
- Taking action for themselves
- Ensure library role-memory organization



www.lockss.org

What Is It Preserving?

- Authoritative versions of heterogeneous content
 - Journals, books, blogs, web sites, scanned files, audio, video ...
- All Web formats and genres
 - Animations, datasets, moving images, still images, software, sound, text ...




www.lockss.org

To Participate...

INSTALLING LOCKSS

Please read this page carefully before bringing your LOCKSS box on line. The page will tell you what you need to know before downloading and installing the software, and will lead you to instructions for doing so.

Bringing up a LOCKSS box is easy and does not require special technical expertise. The LOCKSS box works like an appliance (for example, a TIVO box is an appliance). Librarians have been bringing online LOCKSS boxes for eight years. The first time it may take you as little as an hour -- or it may take you a half day. If it takes longer than this -- contact lockss-support@lockss.org, something is wrong. Once your LOCKSS box is online it requires very little maintenance, about 15 minutes to 1 hour per month.

Watch this YouTube Video of how to bring a LOCKSS box online <http://www.youtube.com/watch?v=0wdcnXrQkaI> 

The LOCKSS software turns a PC into a digital preservation appliance. An appliance does one thing, and because it only does one thing it can be easy to use. The key to making the LOCKSS appliance easy to use is that it boots and runs from a CD and stores all its configuration information on media such as a floppy disk that is write-locked while the system is running. There is no software of any kind, even an operating system, on the PC's disk. The disk is used only to store preserved content.

The hardest things for a system administrator to do correctly are to install, configure and upgrade software so



www.lockss.org

NAVIGATION

- [Home](#)
- [Participating Libraries](#)
- [How It Works](#)
- [Publishers and Titles](#)
- [For Libraries](#)
- [For Publishers](#)
- [LOCKSS Alliance](#)
- [News](#)
- [CLOCKSS](#)
- [Talks](#)
- [Publications](#)
- [Installing LOCKSS](#)

How LOCKSS Works

- Publisher gives permission
 - By putting a Creative Commons license
 - Or a LOCKSS permission statement online
- Publishers include
 - Bloggers!



www.lockss.org

How LOCKSS Works





dshr's blog

897

I'M DAVID ROSENTHAL, AND THIS IS A PLACE TO DISCUSS THE WORK I'M DOING IN DIGITAL PRESERVATION.

monday, July 13, 2009

Spring CNI Plenary: The Video

CNI has now posted the video of Cliff Lynch's introduction, my plenary presentation, and the questions.

How Are We Ensuring the Longevity of Digital Documents?

from CNI Video Editor



blog rules



Posts and comments are copyright of their respective authors who, by posting or commenting, license their work under a Creative Commons Attribution-Share Alike 3.0 United States License. Off-topic or unsuitable comments will be deleted.

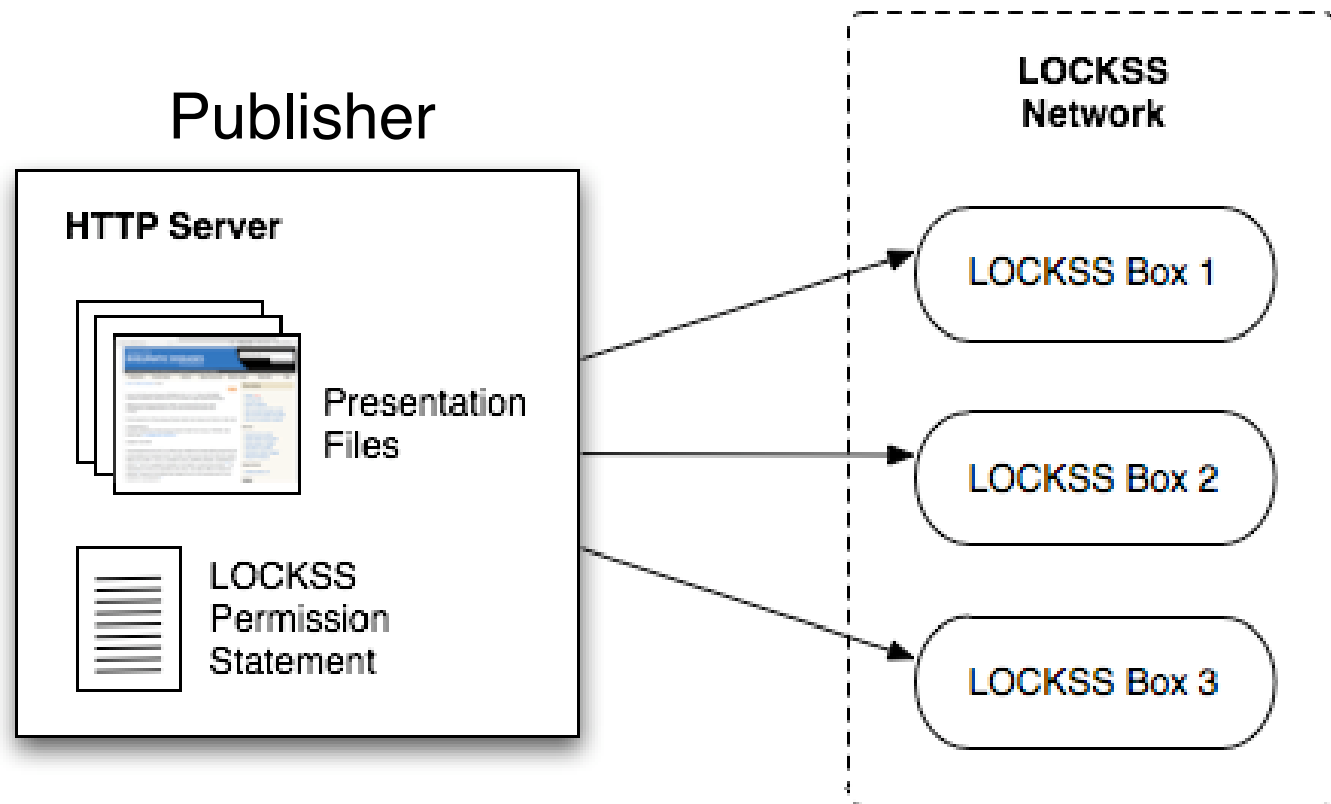
dshr



DSHR in ANWR

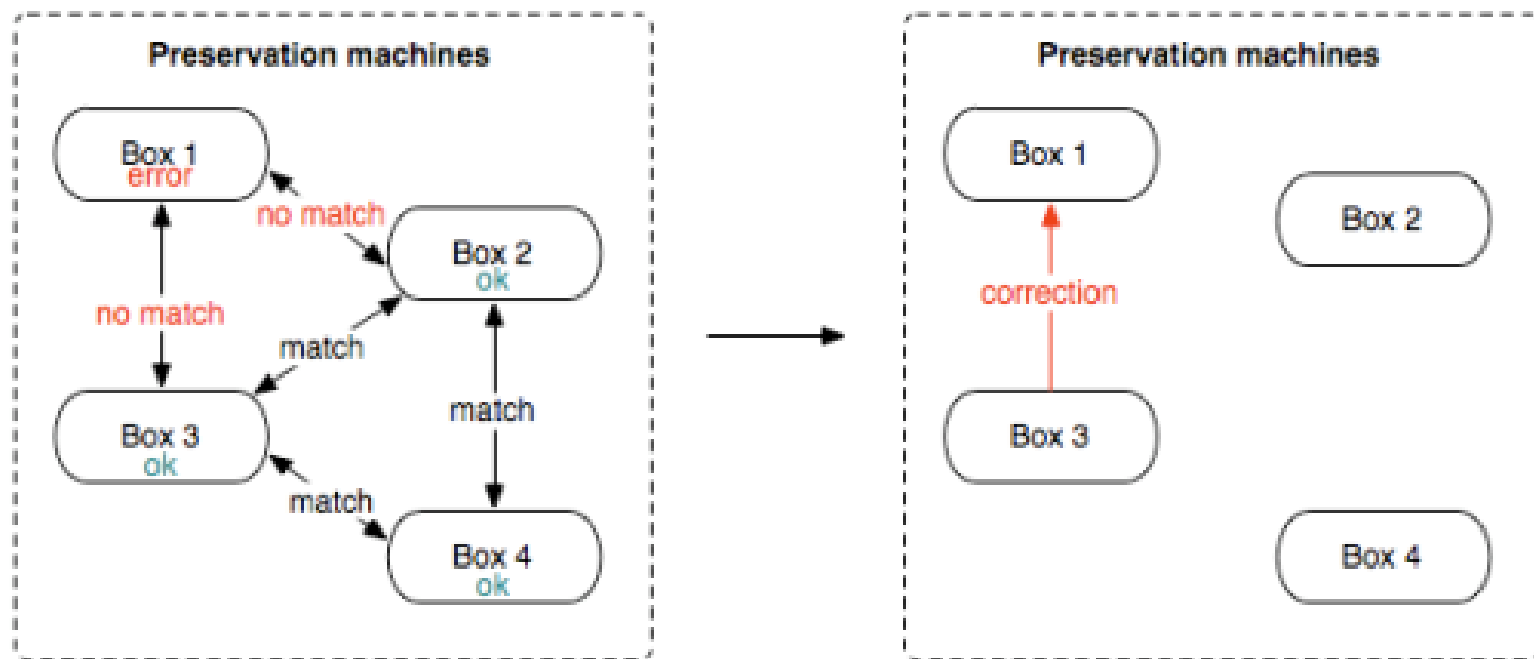
www.lockss.org

How LOCKSS Works



www.lockss.org

How LOCKSS Works



www.lockss.org



Daemon Status

beta2.lockss.org at 17:53:38 03/21/07, up 22h15m17s

Archival Units

Archival Units

1055 Archival Units
11 awaiting recalc

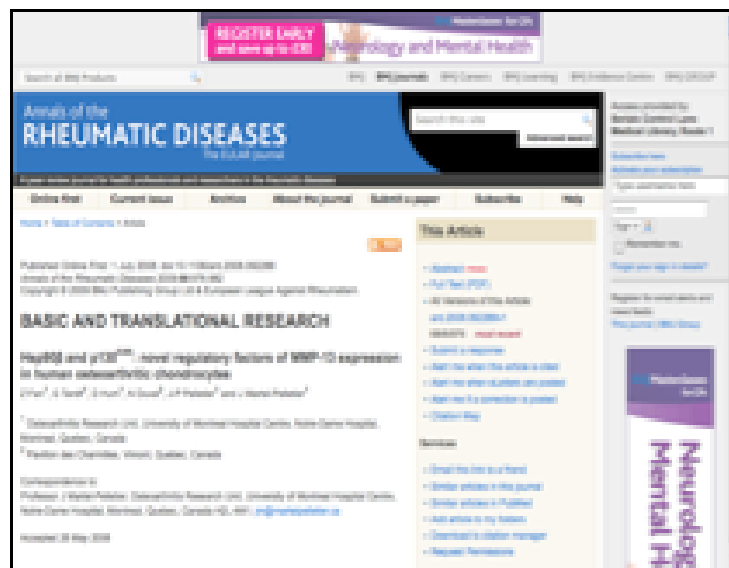
Volume	Content Size ¹	Disk Usage (MB) ¹	Peers	Polls	Status ²	Last Poll
Academic Psychiatry Volume 30	93,400,185	103.9	peers	0	Waiting for Poll Agreement	03/21/07 00:14:49 10/22/06
Accounting History Volume 5	3,125,086	5.4	peers	0	100.00% Agreement	02:30:36 02/12/07
Acta Sociologica Volume 48	7,992,857	11.4	peers	0	86.93% Agreement	00:05:26 02/14/07
Action Research Volume 3	6,443,633	9.6	peers	0	95.77% Agreement	12:47:23 02/03/07
Active Learning in Higher Education Volume 6	5,146,272	7.7	peers	0	Waiting for Poll Agreement	12:03:57 10/28/06
Adaptive Behavior Volume 13	13,264,401	23.4	peers	0	99.39% Agreement	17:11:49 12/23/06



www.lockss.org

How LOCKSS Works

Article provided by the publisher



Article provided by LOCKSS



www.lockss.org

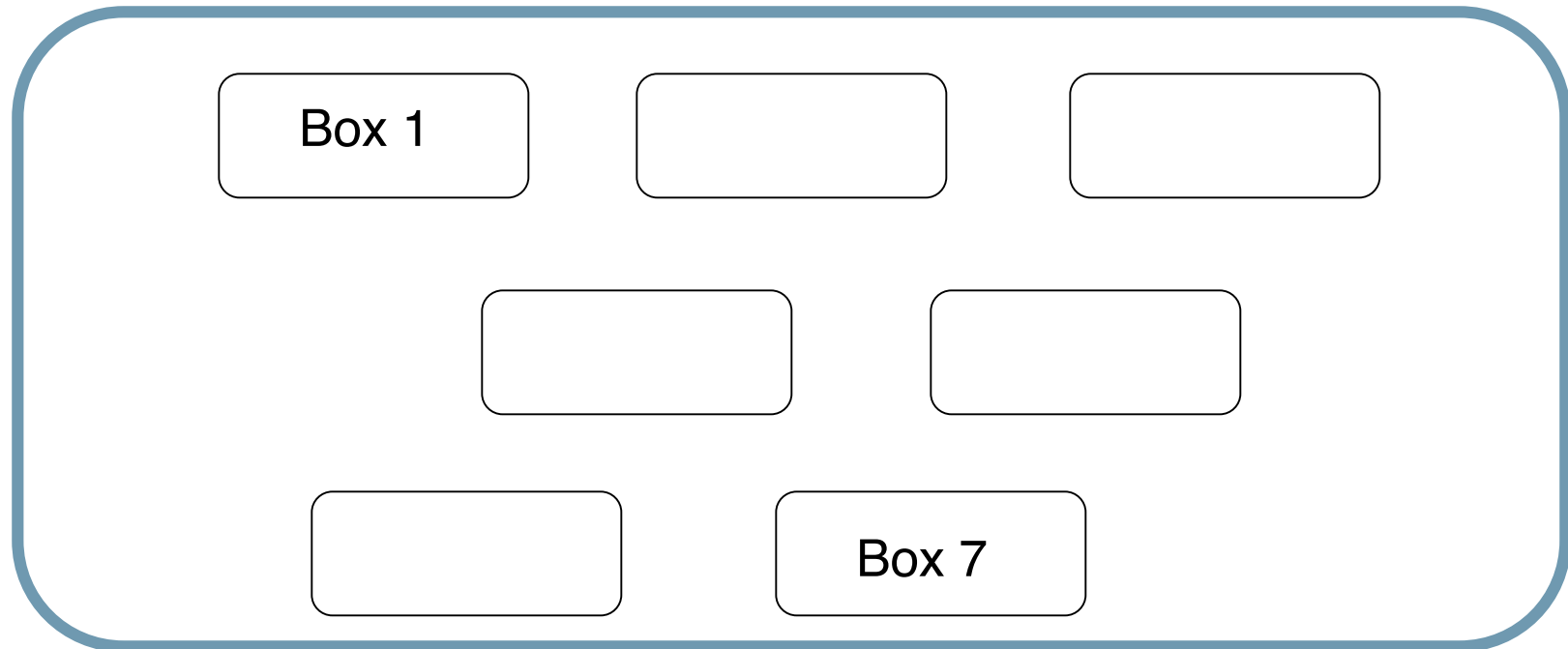
Two Kinds Of LOCKSS Networks

- **Public** LOCKSS network
- **Private** LOCKSS networks
 - Communities share preservation responsibilities
 - You preserve my stuff, I'll preserve yours
 - Arizona State Library and Archives - PeDALS
 - Data-PASS - social science data sets
 - U.S. Federal Government Documents



www.lockss.org

How LOCKSS Works



www.lockss.org

Introduction

- This talk is in 3 parts
 - About LOCKSS
 - What is it?
 - How does it work?
 - Decentralized preservation
 - Authentication
 - Threats to digital content
 - Making preservation easier



www.lockss.org

Centralized Collection & Preservation

- Hub and spoke model
 - Change at hub propagates along spokes
 - Single point of corruption
- Centralized authentication
 - Single point of subversion
 - Only publisher can sign web pages
 - Too expensive for blogs and dynamic content
 - Public key infrastructure lacking



www.lockss.org

Decentralized Collection & Preservation

- Independent collection, independent replicas
 - Very hard to change all copies w/out being caught
 - ACM research award for showing how to make this true
- Distributed ‘authentication’
 - Compare multiple independent collections
 - If consensus, it’s what publisher meant, else:
 - Still changing? (e.g. blog comments, stop after a while)
 - Defective?



www.lockss.org

Threat Model

- What are you preserving stuff from?
- Many preservation models ignore threats known to cause data loss

<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>



www.lockss.org

Threat Model

- Media failure
 - Hardware failure
 - Software failure
 - Network failure
 - Obsolescence
 - Natural Disaster
- But....



www.lockss.org

Threat Model

- Media failure
- Hardware failure
- Software failure
- Network failure
- Obsolescence
- Natural Disaster
- Operator error
- External Attack
- Insider Attack
- Economic Failure
- Organization Failure



www.lockss.org

Introduction

- This talk is in 3 parts
 - About LOCKSS
 - What is it?
 - How does it work?
 - Decentralized preservation
 - Authentication
 - Threats to digital content
 - Making preservation easier



www.lockss.org

Making Preservation Easier

- Use Creative Commons (CC) license
 - Permits archiving through time
 - By any interested institution
 - Facilitates archive interoperability
- You may not think it worth preserving
 - CC allows others to disagree
 - Without having your lawyers & their lawyers talk
 - 1 lawyer-hour > 5TB of storage.



www.lockss.org

Making Preservation Easier

- Use meaningful URLs
 - Canonical, hierarchical
 - Don't use session IDs in links, put in cookies
 - Blogger does this right
- Dynamic elements add complexity
 - What canonical content must be preserved, ads?
 - Use semantic markup & style sheets
- Don't put content behind a fee wall



www.lockss.org

Questions Welcome!

Thank you.



www.lockss.org