What Is It?

Open Source    Peer to Peer    Decentralized
Digital Preservation Infrastructure
2 ASSUMPTIONS
"Libraries are directly and immediately involved in the conflict which divides our world, and for two reasons. First, because they are essential to the functioning of a democratic society. Second, because the contemporary conflict touches the integrity of scholarship, the freedom of the mind, and even the survival of culture, and libraries are the great tools of scholarship, the great repositories of culture, and the great symbols of the freedom of the mind."

--Franklin Delano Roosevelt
A {research} library contains an in-depth collection in a particular subject field ... or in-depth collections in several subject fields ...
The Web

Accidentally changed the business relationship between librarians and publishers.

With rare exceptions, libraries provide access to web materials and do not take physical custody of the content.

The role libraries have played in society for hundreds of years as trusted keepers of information and culture was disrupted.

Our digital cultural and intellectual heritage is at risk.
Archival Assets
2 Key Questions

From this moment on,

• who will have custody
• who will control and govern societies archival assets?
Community Governance
Community Custody
To start: 12 Publishers and 7 libraries
Building a distributed, validated, comprehensive archive
Held on behalf of the community
Global access to content post trigger

HANDOUT
Libraries are using LOCKSS to build libraries; to build & preserve e-collections

• Subscription materials: eliminating perpetual payment for perpetual access
• Freely available materials: truly and immediately at risk
The Community Is Keeping LOTS OF STUFF SAFE

Content that meets these criteria
1. Delivered via the web, http://
2. Has a stable URL structure
3. Has an authoritative version

FORMAT AGNOSTIC -- ALL FILE TYPES
How Does It Work?
Two Requirements

Library installs a LOCKSS Box
www.lockss.org
[install]

Publisher gives permission
## Archive of 2003 Online Issues:

<table>
<thead>
<tr>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>January</strong>; 51 (1): 1 - 102</td>
<td><strong>February</strong>; 51 (2): 103 - 267</td>
<td><strong>March</strong>; 51 (3): 271 - 404</td>
</tr>
<tr>
<td>April</td>
<td>May</td>
<td>June</td>
</tr>
<tr>
<td><strong>April</strong>; 51 (4): 407 - 554</td>
<td><strong>May</strong>; 51 (5): 555 - 696</td>
<td><strong>June</strong>; 51 (6): 697 - 852</td>
</tr>
<tr>
<td>July</td>
<td>August</td>
<td>September</td>
</tr>
<tr>
<td><strong>July</strong>; 51 (7): 853 - 980</td>
<td><strong>August</strong>; 51 (8): 981 - 1112</td>
<td><strong>September</strong>; 51 (9): 1113 - 1248</td>
</tr>
<tr>
<td>October</td>
<td>November</td>
<td>December</td>
</tr>
<tr>
<td><strong>October</strong>; 51 (10): 1249 - 1391</td>
<td><strong>November</strong>; 51 (11): 1393 - 1574</td>
<td><strong>December</strong>; 51 (12): 1575 - 1712</td>
</tr>
</tbody>
</table>

**Publisher Gives Permission**

LOCKSS system has permission to collect, preserve, and serve this Archival Unit.
Collection

The diagram illustrates a collection consisting of two titles: Title 1 and Title 2. Each title is connected to a LOCKSS box through arrows labeled "Patron." The LOCKSS boxes are represented by rectangular shapes with a grid pattern.

- Title 1 is connected to a purple LOCKSS box.
- Title 2 is connected to a red LOCKSS box.

The overall structure is labeled as "LOCKSS Boxes."
Preservation

LOCKSS Boxes

Title 1

Title 2

Patron

LOCKSS box

LOCKSS box

LOCKSS Boxes
## Daemon Status

### Archival Units

<table>
<thead>
<tr>
<th>Archival Units</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Volume</strong></td>
<td><strong>Content Size</strong></td>
<td><strong>Disk Usage (MB)</strong></td>
<td><strong>Peers</strong></td>
<td><strong>Polls</strong></td>
<td><strong>Status</strong></td>
<td><strong>Last Poll</strong></td>
<td><strong>Last Crawl</strong></td>
<td><strong>Last TreeWalk</strong></td>
<td></td>
</tr>
<tr>
<td>Abstinence Literary Review 2003</td>
<td>1,254,583</td>
<td>7.2</td>
<td>peers</td>
<td>13</td>
<td>polls</td>
<td>Ok</td>
<td>06:43:51</td>
<td>06:30:53</td>
<td>17:59:52</td>
</tr>
<tr>
<td>American Forestry Journal Volume 94</td>
<td>6,364,079</td>
<td>13.2</td>
<td>peers</td>
<td>22</td>
<td>polls</td>
<td>Ok</td>
<td>16:56:09</td>
<td>19:00:52</td>
<td>08:16:04</td>
</tr>
<tr>
<td>American Historical Review Volume 108</td>
<td>21,788,288</td>
<td>31.3</td>
<td>peers</td>
<td>1</td>
<td>polls</td>
<td>Ok</td>
<td>09:01:39</td>
<td>06:38:38</td>
<td>01:55:43</td>
</tr>
<tr>
<td>American Jewish History Volume 91</td>
<td>17,493,000</td>
<td>24.9</td>
<td>peers</td>
<td>32</td>
<td>polls</td>
<td>Ok</td>
<td>06:08:07</td>
<td>04:22:32</td>
<td>16:03:16</td>
</tr>
<tr>
<td>The American Midland Naturalist Volume 152</td>
<td>9,177,858</td>
<td>16.6</td>
<td>peers</td>
<td>5</td>
<td>polls</td>
<td>Ok</td>
<td>01:52:54</td>
<td>20:44:41</td>
<td>07:21:46</td>
</tr>
<tr>
<td>American Museum Novitates Volume 3460</td>
<td>943983</td>
<td>2.5</td>
<td>peers</td>
<td>31</td>
<td>polls</td>
<td>Ok</td>
<td>08:30:04</td>
<td>21:35:08</td>
<td>06:27:52</td>
</tr>
<tr>
<td>American Zoologist Volume 41</td>
<td>51,975,406</td>
<td>80.3</td>
<td>peers</td>
<td>0</td>
<td>polls</td>
<td>Ok</td>
<td>23:17:16</td>
<td>14:26:17</td>
<td>21:39:47</td>
</tr>
</tbody>
</table>
Prevents the publisher from revoking access rights to back content
Look and Feel to Readers

When content is served to the user from a LOCKSS Box

– Look and feel is as close as possible to what the publisher published
– Preserve content & presentation
CREELEY’S BIRTHDAY DRIVE
By Andrei Codrescu

Bob Creeley’s birthday parties were legendary, but nobody in their right mind would get in a car with Bob for his traditional birthday drive. Not even...

THE CORPSE READS CLASSICS
(THED CORPSE ALWAYS READS
AND RE-READS THE CLASSICS,
WITH POUND’S INUNCTION IN MIND)

Robert Casella knows his Greeks; they guide him; he gets up to Christos, then stops, wisely.

Dennis DiClaudio revisits the case of Hypatia

Kevin McLeian transcribes certain oracles, from Sicily to Jamaica Plains, concerning homosexuality

Frank Eannarino revisits Genesis for a poetic treatise on “the evolution of panspermia.” We are not sure what he’s after, but the sounds pleased us

Julie Keitges dips her lyric pail into the Renaissance and comes up with a noble English serial killer; not the classics, exactly, but we are selling this one to the History Channel

David Schwartz asks “Where Does Midrash Derive?” and then proceeds to answer like a rabbi
Format Migration

“on the fly”
When content is requested
Process is transparent to the reader

D-Lib Magazine, Volume 11, Number 1, January, 2005
http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html
Format Migration “on the fly”

Preserve the original look & feel. Value in increasing amounts of content.

Reduce the cost of ingest. Preserve more material per dollar.

Postpone costs of migration. Leverage time value of money & technology cost curve.

Migrate material upon reader request.
Vastly lower the amount of content processed.

What the readers sees is the result of the best possible technology at time of access.
Audit: Threats Models
Audit: Bit Preservation
Audit
Threat Models

2005 National Science Board report to NARA

……the designers of a digital preservation system need a clear vision of the threats against which they are being asked to protect their system’s contents, and those threats under which it is acceptable for preservation to fail.
Threats

Continuous and Abrupt changes

Technology
- storage media, hardware, software, formats

Money
- here one day, gone the next

Organizations
- shifting priorities, politics, staffs

Natural disasters

Human errors and attacks
Biggest threat to digital preservation is money

Digital bits are dynamic.
If you don’t preserve them you loose them.

Digital preservation has to be a “marginal expense”; you can’t “halt” the preservation process and continue to have collections.
Bit Preservation

Digital bits are dynamic
  • Must be preserved on a dynamic system

Must have lots of copies
  • Copies must be loosely coupled
  • Don’t make Winston Smith’s job easy
Audit

Bit Preservation

Develop low-level system protocols and standards

Build on peer to peer LOCKSS audit and repair protocol

Build on draft LC bit preservation functional requirements
Join Us