LOCKSS at Cornell

DLF Spring 2005 Forum

Bill Kehoe, Cornell University Library
Topics

• System Administration
• Adding a journal—Plug-ins
• Retrieving items from the local system
• Costs
• Another use for LOCKSS
System Administration

• It’s a snap to install
  – First time: couple of hours
  – After that, almost automatic—the install process uses the previous, site-specific configuration
  – watch over it, do what it asks—write protect the floppy
  – ½ to 1 hour—one eye on the process.
Story—ACL Horrors

• Early days—Cornell had no firewall
• ACL put on individual subnets
  – ACL is the list of holes through the firewall
  – A list—specific protocols allowed at specific ports
    • TCP/IP—8080, 8081/ UDP – 5555, 5554
  – Intent: Block Off-campus traffic into Cornell
• Changed the machine IP—no more incoming traffic
  – Bad for the LOCKSS system, but I’m not worried—it repairs itself
  – Central IT folks say it’s not their fault; Tom says it’s not the LOCKSS system
  – Two months of back and forth, not every day, because I don’t want to insult anyone
  – Email from CIT: “Oops.”
Anecdotal plug-in writing experience

• First time—maybe two hours, as I learned how the harvesting rules worked, and a back and forth with Tom
• After that—less than half an hour
• I don’t do it frequently enough to remember what’s required for each element
• The Building a Plug-in tutorial is very helpful.
What’s not so clear about plug-in writing

• What is the atomic unit for the journal—a volume, an issue, a year?
• Should there be a plug-in for each unit?
• Some publishers change directory naming conventions over the years.
• How many plug-ins to build?
Retrieving items from the local system

- The LOCKSS Winter 2005 Card
- http://libdev.library.cornell.edu/lockss
Time Costs

• Time
  – Selector
    • Selection decision—15 minutes to 1 hour
    • Conversation with publisher—30 minutes to 1 hour
  – Techie
    • Evaluate journal for technical feasibility—30 minutes to 1 hour
    • Write plug-in—15 minutes to 1 hour
    • Converse with publisher techie and LOCKSS team—30 minutes
  • Total—2 to 4 ½ hours
### Network Costs

<table>
<thead>
<tr>
<th>Date</th>
<th>Conversations</th>
<th>Traffic (MB)</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-11</td>
<td>3680</td>
<td>470.98</td>
<td>$3.00</td>
</tr>
<tr>
<td>2004-12</td>
<td>3253</td>
<td>1630.97</td>
<td>$3.00</td>
</tr>
<tr>
<td>2005-03</td>
<td>2140</td>
<td>7262.98</td>
<td>$13.43</td>
</tr>
<tr>
<td>My desktop 2005-03</td>
<td>2826</td>
<td>716.35</td>
<td>$3.00</td>
</tr>
</tbody>
</table>
Another Use for LOCKSS

- Ensuring Access to Mathematics Over Time
  [http://www.library.cornell.edu/dlit/EATMOT](http://www.library.cornell.edu/dlit/EATMOT)
- OAIS information model lacks a level of metadata that describes partnerships in collaborative archives.
- EATMOT repository: XML-encoded metadata about the partnership members, rights agreements, storage capabilities, preservation policies.
- LOCKSS for the repository—inexpensive, format-agnostic, self-repairing, no one owns it.