SafeArchive
Audit System Overview

Jonathan Crabtree
Odum Institute UNC Chapel Hill
Data preservation requires replication

LOTS OF COPIES KEEP STUFF SAFE
Daemon 1.51.6 built 28-Sep-11 14:45:30 on build7.lockss.org, Linux RPM 1
Managing copies can be challenging
Facilitating Collaborative Preservation

- **Collaborators** make resource commitments
- **Policy** records commitments, system properties, TRAC policies
- **Content is stored in** The Dataverse Network, other standard repository systems, on a web server, etc.
- **LOCKSS** provides replication, integrity, freshness, harvesting
- **SafeArchive software** provides monitoring, auditing, restoration, and provisioning
Centralize Network Auditing

Collaborating Institutions

Policies

Schema Form

Content

TRAC Input Form

Preservation Network

Discovered Network Data

The SafeArchive System

Audit Schema Manager
- Create Audit Schema
- Edit Audit Schema
- Schema Comparison Tool
- Preservation Enforcer

Network Monitor
- Cache Status Extractor
- Cron Service
- Console

Report Generator
- View Audit Reports
- View Operational Reports
- Manually Refresh Audit
- Generate Historical Reports

LOCKSS Management Tool
(future development)

Comparison

Audit Report

Reports

Operational Reports

Repair or Reprovision
### Network Data Tables

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Show Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>lockss_box_table</td>
<td>Show Table</td>
</tr>
<tr>
<td>archival_unit_status_table</td>
<td>Show Table</td>
</tr>
<tr>
<td>repository_space_table</td>
<td>Show Table</td>
</tr>
<tr>
<td>au_overview_table</td>
<td>Show Table</td>
</tr>
<tr>
<td>crawl_status_table</td>
<td>Show Table</td>
</tr>
<tr>
<td>polls_table</td>
<td>Show Table</td>
</tr>
<tr>
<td>successful_polls_table</td>
<td>Show Table</td>
</tr>
</tbody>
</table>

Data collection as of: 2011-10-18T00:00:33

### Refresh Network-Status Data Tables

To browse the latest network-status data:

1. Click the refresh button below.
2. Wait for the refresh-completion message.
3. Log out and log in the SAAS again.

[Refresh Tables]

### Operation Reports

**By Host**

- View report

**By Archival-unit (AU)**

- View report
<table>
<thead>
<tr>
<th>PinAuId</th>
<th>AuName</th>
<th>AuSizeMax</th>
<th>AuNReplicas</th>
<th>AuNVerifiedReplicas</th>
<th>LastSPollEnd</th>
<th>LastSCrawlEnd</th>
<th>CrawlDuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2341</td>
<td>IQSS Dataverse</td>
<td>95,744.0</td>
<td>6</td>
<td>5</td>
<td>2011-10-19 08:25:12 EDT</td>
<td>2011-10-17 21:29:38 EDT</td>
<td>0019:00:19</td>
</tr>
<tr>
<td>2339</td>
<td>ICPSR Dataverse</td>
<td>299,432.0</td>
<td>5</td>
<td>0</td>
<td>2011-10-18 13:47:49 EDT</td>
<td>2011-10-18 13:47:49 EDT</td>
<td>0019:00:31</td>
</tr>
<tr>
<td>2343</td>
<td>Roper's Data-PASS subset</td>
<td>1,725.0</td>
<td>7</td>
<td>6</td>
<td>2011-10-18 20:02:46 EDT</td>
<td>2011-10-14 15:04:18 EDT</td>
<td>0019:00:04</td>
</tr>
<tr>
<td>2340</td>
<td>Index of /locks /plugins</td>
<td>0.0</td>
<td>7</td>
<td>5</td>
<td>2011-10-18 20:51:16 EDT</td>
<td>2011-10-18 23:55:41 EDT</td>
<td>0019:00:00</td>
</tr>
</tbody>
</table>
### Archival-Unit Overview: List

<table>
<thead>
<tr>
<th>PnAuld</th>
<th>AuName</th>
<th>AuSizeMax</th>
<th>AuNReplicas</th>
<th>AuNVerifiedReplicas</th>
<th>LastSPollEnd</th>
<th>LastSCrawlEnd</th>
<th>CrawlDuration</th>
</tr>
</thead>
</table>

- **PnAuld:** 2342
- **AuName:** Odum Dataverse
- **AuSizeMax:** 62,747.0
- **AuNReplicas:** 7
- **AuNVerifiedReplicas:** 6
- **LastSPollEnd:** 2011-10-19 02:50:57 EDT
- **LastSCrawlEnd:** 2011-10-14 10:28:34 EDT
- **CrawlDuration:** 0019:19:47

---

<table>
<thead>
<tr>
<th>PnAuld</th>
<th>IQSS Dataverse</th>
<th>IQSS SizeMax</th>
<th>IQSS NReplicas</th>
<th>IQSS NVerifiedReplicas</th>
<th>IQSS LastSPollEnd</th>
<th>IQSS LastSCrawlEnd</th>
<th>IQSS CrawlDuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2341</td>
<td>95,744.0</td>
<td>6</td>
<td>5</td>
<td>2011-10-19 08:25:12 EDT</td>
<td>21:29:38 EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2339</td>
<td>299,432.0</td>
<td>5</td>
<td>0</td>
<td>2011-10-18 13:47:49 EDT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2343</td>
<td>1,725.0</td>
<td>7</td>
<td>6</td>
<td>2011-10-18 20:02:46 EDT</td>
<td>15:04:18 EDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2340</td>
<td>0.0</td>
<td>7</td>
<td>5</td>
<td>2011-10-18 20:51:16 EDT</td>
<td>23:55:41 EDT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Defining Policy

- Storage commitment
- Space allocation requirement
- Geographic replication
- Replication
- Verification frequency
- Data update frequency
- TRAC audit policies
### List Audit Schema Instances

<table>
<thead>
<tr>
<th>Schema Instance Name</th>
<th>Last Modified Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>lockss_audit_schema_instance.xml</td>
<td>2011-10-17 10:08:27 EDT</td>
</tr>
</tbody>
</table>

- **Create a New Audit Schema Instance**
- **Back to the Home Page**

### Current Audit Schema Instance

Name: lockss_audit_schema_instance.xml
Last Modified Date: 2011-10-17 10:08:27 EDT

### Audit Schema Instances under Version Control

<table>
<thead>
<tr>
<th>Revision Number</th>
<th>Committed Date</th>
<th>Saved By</th>
<th>Log Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>146</td>
<td>2011-10-17 10:08:27 EDT</td>
<td>curator</td>
<td>Audit Schema Instance:id=1.0.0; Version=asid_89x; date=2011-10-17T10-08-27; saved by curator;</td>
</tr>
<tr>
<td>145</td>
<td>2011-10-17 10:01:25 EDT</td>
<td>curator</td>
<td>Audit Schema Instance:id=1.0.0; Version=asid_f4YU; date=2011-10-17T10-01-24; saved by curator;</td>
</tr>
<tr>
<td>144</td>
<td>2011-10-17 09:58:58 EDT</td>
<td>curator</td>
<td>Audit Schema Instance:id=1.0.0; Version=asid_gx8; date=2011-10-17T09-58-58; saved by curator;</td>
</tr>
<tr>
<td>143</td>
<td>2011-10-17 09:53:48 EDT</td>
<td>curator</td>
<td>Audit Schema Instance:id=1.0.0; Version=asid_3FyS; date=2011-10-17T09-53-48; saved by curator;</td>
</tr>
<tr>
<td>142</td>
<td>2011-10-17 09:52:42 EDT</td>
<td>curator</td>
<td>Audit Schema Instance:id=1.0.0; Version=asid_jn7; date=2011-10-17T09-52-42; saved by curator;</td>
</tr>
</tbody>
</table>

(1 of 9)
### Edit Audit Schema Instance

#### About the PLN

Question 1: What is the name of the private LOCKSS network (PLN) that you would like the SAFE system to audit?  
Question 2: What is the email address of the PLN administrator?

#### About PLN member Hosts

Question 3: How much storage disk space do you wish to commit on each server for the SAFE system to use (in gigabytes)?

<table>
<thead>
<tr>
<th>host Name</th>
<th>host IP Address</th>
<th>Space Committed (Gb)</th>
<th>Geographic Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>lockss-3.ropercenter.uconn.edu</td>
<td>137.99.36.160</td>
<td>1740</td>
<td>Connecticut</td>
</tr>
<tr>
<td>lockss.hmdc.harvard.edu</td>
<td>140.247.115.220</td>
<td>39</td>
<td>Massachusetts</td>
</tr>
<tr>
<td>lockss-0.icpsr.umich.edu</td>
<td>241.211.146.29</td>
<td>120</td>
<td>Michigan</td>
</tr>
<tr>
<td>lockss-1.icpsr.umich.edu</td>
<td>141.211.146.52</td>
<td>1843</td>
<td>Michigan</td>
</tr>
<tr>
<td>lockss-2.icpsr.umich.edu</td>
<td>141.211.146.53</td>
<td>1843</td>
<td>Michigan</td>
</tr>
</tbody>
</table>

The above table shows discovered servers only. To add a missing server or update a host's attributes, click the following link:

**Edit the Host List**

#### About Archival Units

Note: Before preservation content can be audited, it must be placed on a Host as an Archival unit or AU. If your content is not listed below in an AU, click the following link to add it to a LOCKSS host:

**Edit the Host List**

Question 4: The following AUs have been found on the specified hosts. Please check the archival units you would like to audit and answer the following questions in respective columns of the table below:
The above table shows discovered servers only. To add a missing server or update a host’s attributes, click the following link:

Edit the Host List

About Archival Units

Note: Before preservation content can be audited, it must be placed on a Host as an Archival unit or AU. If your content is not listed below in an AU, click the following link to add it to a LOCKSS host:

Edit the Host List

Question 4: The following AUs have been found on the specified hosts. Please check the archival units you would like to audit and answer the following questions in respective columns of the table below:

4.1 How frequently should this archival unit AU be re-crawled (in days)? Re-crawling adds new content to the AUs as changes occur keeping your preservation copies current this is the Update Frequency.

4.2 What is the maximum size that you expect this archival unit AU to be (in gigabytes)? The default value is the current size. Please add your growth expectations. This is your storage required.

Question 5: How many preservation copies of each collection do you want to maintain? This will be the Number of Replicates.

Question 5a: How many regions do you want to maintain a replica of a collection? This will be the value of Geographic Redundancy.

Question 6: Archival Units AUs will be tested on a regular basis for problems to ensure your content is not corrupt. What is the maximum time you would allow between these verifications, in days? This is your Verification Frequency.

Question 7: What is the maximum amount of time that the preservation system should take to ingest/crawl your content (in days)? LOCKSS uses crawling to add your content to an AU. The maximum number of days this should take is your Update Duration.

<table>
<thead>
<tr>
<th>Archival Unit Name</th>
<th>Update Frequency</th>
<th>Storage Required (Gb)</th>
<th>Number of Replicates</th>
<th>Geographic Redundancy</th>
<th>Verification Frequency</th>
<th>Update Duration</th>
<th>Subject</th>
<th>Owner Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICPSR Dataverse</td>
<td></td>
<td>500</td>
<td>3</td>
<td>2</td>
<td>21</td>
<td>21</td>
<td>Social Sciences</td>
<td>UMich-ICPSR</td>
</tr>
<tr>
<td>Index of /lockss/plugins</td>
<td></td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>21</td>
<td>21</td>
<td>Social Sciences</td>
<td>HU-IQSS</td>
</tr>
<tr>
<td>IQSS Dataverse</td>
<td></td>
<td>500</td>
<td>3</td>
<td>2</td>
<td>21</td>
<td>21</td>
<td>Social Sciences</td>
<td>HU-IQSS</td>
</tr>
<tr>
<td>Odum Dataverse</td>
<td></td>
<td>300</td>
<td>3</td>
<td>2</td>
<td>21</td>
<td>21</td>
<td>Social Sciences</td>
<td>UNC-Odum Institute</td>
</tr>
<tr>
<td>Roper's Data-PASS subset</td>
<td></td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>21</td>
<td>21</td>
<td>Social Sciences</td>
<td>UConn-Roper</td>
</tr>
</tbody>
</table>

Q4.1 Q4.2 Q5 Q5a Q6 Q7

To update or correct an AU’s subject or ownership information, click the following link:

Edit the Archival-Unit List
SafeArchive - Safe Archive Audit System (SAAS)

Audit Schema Instance List
Edit TRAC Audit Checklist

Audit Reporting  Management Tools  Configuration Console  Preservation Enforcer  Support

curator - Logout

Audit Schema Instance List
Edit TRAC Audit Checklist

SAFE PLN: Poll Agreement Results by AU

<table>
<thead>
<tr>
<th>AU</th>
<th>Latest-3</th>
<th>Latest-2</th>
<th>Latest-1</th>
<th>Latest Poll</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICPSR Dataverse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQSS Dataverse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index of /lockss/plugins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odum Dataverse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As of 2011-10-19T00:00:00-23

Superscript: 100%  less than 100%  Cancelled  In Progress

Experimental Showcase 1
Archival Units: Available (dark blue) and Verified (light blue)

Experimental Showcase 2
Verified Odum-Dataverse Replicas by Region

Map of the United States with color-coded regions.
### Section A

<table>
<thead>
<tr>
<th>Id</th>
<th>Criterion</th>
<th>Evidence (Documents) Examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1.1</td>
<td>Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.</td>
<td><a href="http://www.icpsr.umich.edu/icpsrweb/DATAPASS/about.jsp#overview">http://www.icpsr.umich.edu/icpsrweb/DATAPASS/about.jsp#overview</a>, Data-PASS Articles of Collaboration: <a href="http://www.icpsr.umich.edu/files/DATAPASS/pdf/collaboration.pdf">http://www.icpsr.umich.edu/files/DATAPASS/pdf/collaboration.pdf</a></td>
</tr>
<tr>
<td>A1.2</td>
<td>Repository has an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.</td>
<td>Data-PASS Articles of Collaboration</td>
</tr>
<tr>
<td>A2.1</td>
<td>Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfill these duties.</td>
<td>Data-PASS Articles of Collaboration, [Odum Staff Page and Bios](<a href="http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp?partner's">http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp?partner's</a> Staff Pages)</td>
</tr>
<tr>
<td>A2.2</td>
<td>Repository has the appropriate number of staff to support all functions and services.</td>
<td>Data-PASS Publications and Presentations: [<a href="http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp?partner's">http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp?partner's</a> Staff Pages](<a href="http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp?partner's">http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp?partner's</a> Staff Pages), Data-PASS Publications and Presentations, <a href="http://www.icpsr.umich.edu/files/DATAPASS/pdf/collaboration.pdf">Odum Internal Courses</a></td>
</tr>
</tbody>
</table>
Generate Audit Reports

- PLN Replication status report
- TRAC audit information
- Individual institutional reports
- Manual & automatic reports
- Audit reports stored and versioned on regular basis
### SAFE PLN: Poll Agreement Results by AU

<table>
<thead>
<tr>
<th>AU</th>
<th>Latest-3</th>
<th>Latest-2</th>
<th>Latest-1</th>
<th>Latest Poll</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICPSR Dataverse</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>IQSS Dataverse</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>Index of /lockss/plugins</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>Odum Dataverse</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
</tbody>
</table>

As of 2011-10-18T00:00:33

100% | less than 100% | Cancelled | In Progress

### Experimental Showcase 1

Archival Units: Available (dark blue) and Verified (light blue)

### Experimental Showcase 2

Verified Odum-Dataverse Replicas by Region

![Map of the United States with color-coded regions indicating replica counts.](Image)
TRAC Audit Report

Generate TRAC Audit Report

To generate a new TRAC audit-report, click the following button

Generate a new TRAC Audit Report

Current TRAC Audit Report: Summary and Full Report by BIRT Viewer

To view the latest TRAC audit report, click one of the following links:

TRAC Audit Report: Summary

Full TRAC audit report by BIRT viewer
**Trusted Repositories Audit & Certification (TRAC) Report**

**Oct 19, 2011 4:20 PM**

### TRAC Audit Summary

<table>
<thead>
<tr>
<th>PLN Group Name</th>
<th>safe</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAC version</td>
<td>1.0</td>
</tr>
<tr>
<td>TRAC Report data generation date</td>
<td>Tue Oct 18 13:26:18 EDT 2011</td>
</tr>
</tbody>
</table>

**Complete**

- Section A: 24 out of 24 entered
- Section B: 44 out of 44 entered
- Section C: 15 out of 16 entered

### TRAC Audit Checklist

<table>
<thead>
<tr>
<th>Section A</th>
<th>Incomplete</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1.1</td>
<td></td>
</tr>
<tr>
<td>A1.2</td>
<td></td>
</tr>
<tr>
<td>A2.1</td>
<td></td>
</tr>
<tr>
<td>A2.2</td>
<td></td>
</tr>
<tr>
<td>A2.3</td>
<td></td>
</tr>
<tr>
<td>A3.1</td>
<td></td>
</tr>
</tbody>
</table>

**Data-PASS About Pages:**

- [http://www.icpsr.umich.edu/icpsrweb/DATAPASS/about.jsp#overview](http://www.icpsr.umich.edu/icpsrweb/DATAPASS/about.jsp#overview)
- [http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp, partner’s Staff Pages](http://www.icpsr.umich.edu/icpsrweb/DATAPASS/presentations.jsp)
Update Audit Report Data

Network-daemon-status data were last updated at: 2011-10-19T00:00:23
Audit-report data were last updated at: 2011-10-18T20:09:52
The state of required data tables: Network data are ready for audit-reporting

To generate a new audit-report, click the following button to update audit-report data

[Update Audit Report Data]

Current Audit Report

To view the latest audit report, click one of the following links:
view the current report without filters
view the current report with filters

Archived Audit Reports

<table>
<thead>
<tr>
<th>Revision Number</th>
<th>Committed Date</th>
<th>Log Message</th>
<th>download</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>2011-10-16 01:00:02 EDT</td>
<td>audit report.xml: generation date=2011-10-16T01-00-00; audit schema instance: version=; saved by auditreportC();</td>
<td>download</td>
</tr>
<tr>
<td>132</td>
<td>2011-10-09 01:00:06 EDT</td>
<td>audit report.xml: generation date=2011-10-09T01-00-00; audit schema instance: version=; saved by auditreportC();</td>
<td>download</td>
</tr>
<tr>
<td>129</td>
<td>2011-10-02 01:00:03 EDT</td>
<td>audit report.xml: generation date=2011-10-02T01-00-00; audit schema instance: version=; saved by</td>
<td>download</td>
</tr>
</tbody>
</table>
Audit Summary Report
Report Generation Date: Oct 19, 2011 10:32 AM

Network ID: safe
Reporting Host: dev.irss.unc.edu
Audit Data Date: Tue Oct 18 2011 20:09:52 GMT-0400 (EDT)

### Preservation Network Summary

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Up Time for LOCKSS Hosts</td>
<td>11d:2h:26m:1s</td>
</tr>
<tr>
<td>Number of Hosts in the PLN</td>
<td>7</td>
</tr>
<tr>
<td>Number of Hosts NOT Reporting</td>
<td>0</td>
</tr>
<tr>
<td>Number of Unique AUs in the PLN</td>
<td>5</td>
</tr>
<tr>
<td>Total Number of AU Replicates</td>
<td>32</td>
</tr>
<tr>
<td>Number of AUs NOT Policy Compliant</td>
<td>2</td>
</tr>
<tr>
<td>Total Disk Space in the PLN</td>
<td>21.555 T</td>
</tr>
<tr>
<td>Total Disk Space in Use</td>
<td>2.651 T</td>
</tr>
<tr>
<td>Total Disk Space Free</td>
<td>18.904 T</td>
</tr>
<tr>
<td>Number of Hosts NOT Meeting Storage Commitments</td>
<td>1</td>
</tr>
</tbody>
</table>

### Archival Unit (AU) Summary

**Click here to view Archival Unit Detail Report.**

#### AU Disk Size (estimated)

<table>
<thead>
<tr>
<th>AU Name</th>
<th>Policy</th>
<th>Actual</th>
<th>Audit Result</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQSS Dataverse</td>
<td>500,000 G</td>
<td>93,500 G</td>
<td>Pass</td>
<td>-</td>
</tr>
<tr>
<td>Index of /lockss/plugins</td>
<td>1,000 G</td>
<td>38,055 K</td>
<td>Pass</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Number of Verified AU Replicates

<table>
<thead>
<tr>
<th>AU Name</th>
<th>Policy</th>
<th>Actual</th>
<th>Audit Result</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQSS Dataverse</td>
<td>3</td>
<td>5</td>
<td>Pass</td>
<td>-</td>
</tr>
<tr>
<td>Index of /lockss/plugins</td>
<td>3</td>
<td>6</td>
<td>Pass</td>
<td>-</td>
</tr>
</tbody>
</table>
Archival Unit Detail Report

Network ID: safe
Reporting Host: devins.unc.edu
Report Data Date: Tue Oct 18 23:58:14 EDT 2011

Archival Unit (AU) Overview

<table>
<thead>
<tr>
<th>AU Name</th>
<th>Max AU Replicate Disk Size</th>
<th>Max Allowable Disk Usage</th>
<th>Number of Replicates</th>
<th>Number of Verified Replicates</th>
<th>Last Successful Poll End Date</th>
<th>Last Successful Crawl End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICPSR Dataverse</td>
<td>252.414 G</td>
<td>323.422 G</td>
<td>5</td>
<td>0</td>
<td>Oct 18, 2011 1:47 PM</td>
<td>Oct 18, 2011 1:47 PM</td>
</tr>
<tr>
<td>Roper's Data PASS subset</td>
<td>1.885 G</td>
<td>1.882 G</td>
<td>7</td>
<td>5</td>
<td>Oct 18, 2011 8:02 AM</td>
<td>Oct 14, 2011 3:04 PM</td>
</tr>
</tbody>
</table>

All Replicate Locations and Details

<table>
<thead>
<tr>
<th>Host Name</th>
<th>AU Replicate Disk Size</th>
<th>Allowable Disk Usage</th>
<th>Number of Successful Polls</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>lockss-3 ropercenter.unc.edu</td>
<td>291.332 G</td>
<td>303.990 G</td>
<td>2</td>
<td>Crawling</td>
</tr>
<tr>
<td>lockss-1 icps.umich.edu</td>
<td>292.299 G</td>
<td>304.196 G</td>
<td>1</td>
<td>99.92% Agreement</td>
</tr>
<tr>
<td>lockss-2.icps.umich.edu</td>
<td>292.414 G</td>
<td>323.422 G</td>
<td>1</td>
<td>99.92% Agreement</td>
</tr>
<tr>
<td>fong.iris.unc.edu</td>
<td>292.222 G</td>
<td>304.281 G</td>
<td>0</td>
<td>98.37% Agreement</td>
</tr>
<tr>
<td>host.iris.unc.edu</td>
<td>38.055 K</td>
<td>102.400 K</td>
<td>13</td>
<td>100.00% Agreement</td>
</tr>
</tbody>
</table>

IQSS Dataverse

<table>
<thead>
<tr>
<th>Host Name</th>
<th>AU Replicate Disk Size</th>
<th>Allowable Disk Usage</th>
<th>Number of Successful Polls</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>lockss-3 ropercenter.unc.edu</td>
<td>38.055 K</td>
<td>102.400 K</td>
<td>13</td>
<td>100.00% Agreement</td>
</tr>
</tbody>
</table>
Update Audit Report Data

Network-daemon-status data were last updated at: 2011-10-19T00:00-23
Audit-report data were last updated at: 2011-10-18T20:09-52
The state of required data tables: Network data are ready for audit-reporting

To generate a new audit-report, click the following button to update audit-report data
Update Audit Report Data

Current Audit Report

To view the latest audit report, click one of the following links:
view the current report without filters
view the current report with filters

Archived Audit Reports

<table>
<thead>
<tr>
<th>Revision Number</th>
<th>Committed Date</th>
<th>Log Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>2011-10-16 01:00:02 EDT</td>
<td>audit report xml: generation date=2011-10-16T01-00-00; audit schema instance: version=; saved by auditreportCJ; download</td>
</tr>
<tr>
<td>132</td>
<td>2011-10-09 01:00:06 EDT</td>
<td>audit report xml: generation date=2011-10-09T01-00-00; audit schema instance: version=; saved by auditreportCJ; download</td>
</tr>
<tr>
<td>129</td>
<td>2011-10-02 01:00:03 EDT</td>
<td>audit report xml: generation date=2011-10-02T01-00-00; audit schema instance: version=; saved by auditreportCJ; download</td>
</tr>
</tbody>
</table>
Update Audit Report Data

Network-daemon-status data were last updated at:
Audit-report data were last updated at:
The state of required data tables:

To generate a new audit-report, click the following button to update audit-report data:

Current Audit Report

To view the latest audit report, click one of the following links:
view the current report without filters
view the current report with filters

Archived Audit Reports

<table>
<thead>
<tr>
<th>Revision Number</th>
<th>Committed Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>2011-10-16 01:00:02 EDT</td>
</tr>
<tr>
<td>132</td>
<td>2011-10-09 01:00:06 EDT</td>
</tr>
<tr>
<td>129</td>
<td>2011-10-02 01:00:03 EDT</td>
</tr>
<tr>
<td>126</td>
<td>2011-09-25 01:00:02 EDT</td>
</tr>
</tbody>
</table>

Opening lockss_audit_report_revision_137.xml

You have chosen to open

- lockss_audit_report_revision_137.xml
  which is a XML Document
  from: http://dev.irsc.unc.edu/888

What should Firefox do with this file?
- [ ] Open with: XML Editor (default)
- [ ] Save File
- [ ] Do this automatically for files like this from now on.

[OK] [Cancel]
<xml version="1.0" encoding="UTF-8" ?>
<compare xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="lockss_audit_report.xsd">
  <whatsthis>
    <filename>/usr/safe/data/audit_output_files/lockss_audit_report.xml</filename>
    <date>2011-10-16 01:00:00</date>
    <audit_id>1.0.0</audit_id>
    </whatsthis>
  <hosts>
    - <host host="141.211.146.29" seq_num="1">
      <test>
        <test>
          <schema_xpath>/SSP/hosts/host/@host_id</schema_xpath>
          <schema_value>141.211.146.29</schema_value>
          <db_column>ip_address</db_column>
          <db_value>141.211.146.29</db_value>
          <match>true</match>
          <note />
        </test>
      </test>
      <test>
        <schema_xpath>/SSP/hosts/host Capacities /storageAvailable/@max_size</schema_xpath>
        <schema_value>128849018880</schema_value>
        <db_column>repo_size</db_column>
        <db_value>128849018880</db_value>
        <match>true</match>
        <note />
      </test>
      <test>
        <schema_xpath>/SSP/hosts/host Identity/name</schema_xpath>
        <schema_value>lockss-0.icpsr.umich.edu</schema_value>
        <db_column>host_name</db_column>
        <db_value>lockss-0.icpsr.umich.edu</db_value>
        <match>true</match>
        <note />
      </test>
      <test>
        <schema_xpath>/SSP/network/@network_id</schema_xpath>
        <schema_value>safe</schema_value>
        <db_column>group_name</db_column>
        <db_value>safe</db_value>
        <match>true</match>
        <note />
      </test>
    </host>
    - <host host="141.211.146.52" seq_num="2">
      <test>
        <schema_xpath>/SSP/hosts/host/@host_id</schema_xpath>
        <schema_value>141.211.146.52</schema_value>
        <db_column>ip_address</db_column>
        <db_value>141.211.146.52</db_value>
        <match>true</match>
        <note />
      </test>
    </host>
  </hosts>
</compare>
LOCKSS Lots of Copies Keep Stuff Safe

- Widely used in library community
- Self-contained OSS replication system, low maintenance, inexpensive
- Harvests resources via web-crawling (mostly)…
- Maintains copies through secure p2p protocol
- Zero trust & self repairing

What does SafeArchive Add?

- Auditing – easily monitor number of copies of content in network
- Collaboration – coordinate across partners, monitor resource commitments
- Provisioning – ensure sufficient copies and distribution
- Restoration guarantees
- Integrates with The Dataverse Network ® (and other repositories through OAI-PMH + DDI)
Future Possibilities

- Support additional audit standards
  - Data Seal of Approval
  - ISO 16363
- Simplify SAFE audit standards interface
  - Single interview style questionnaire
  - Automatic crosswalk to all supported audit standards
- Enhance qualitative audit compliance
  - Interactive audit agency tools
  - Multiple levels of audit compliance
- Support additional replication networks
  - iRODS
Collaborators*


Research Support

Thanks to the, IMLS (LG-05-09-0041-09), Library of Congress (PA#NDP03-1), the National Science Foundation (DMS-0835500, SES 0112072), the Harvard University Library, the Institute for Quantitative Social Science, the Harvard-MIT Data Center, and the Murray Research Archive.

* And co-conspirators
Related Work

- Reprints available from: http://futurelib.org


Questions

- **Website**
  - [www.safearchive.org](http://www.safearchive.org)

- **Sourceforge**

- **Contacts**
  - Micah.Altman@gmail.com
  - Jonathan_Crabtree@unc.edu