SFX Integration Guide

LOCKSS

SFX Integration Guide

Table of Contents

Customizing Your LOCKSS Box....3
Enabling the Content Server....3
Configuring Content Server Access Control....4
Using Apache as a Front End to your LOCKSS Content Server....5
Customizing your SFX Instance....7
Configuring the LOCKSS Target and Target Service in SFX....7
Activating Object Portfolios....7
Activate all available portfolios. 7
Activate only configured portfolios. 8
Activate only collected portfolios. 8
Activate only discontinued portfolios. 9
Using DataLoader to Activate Object Portfolios....9
Troubleshooting....10

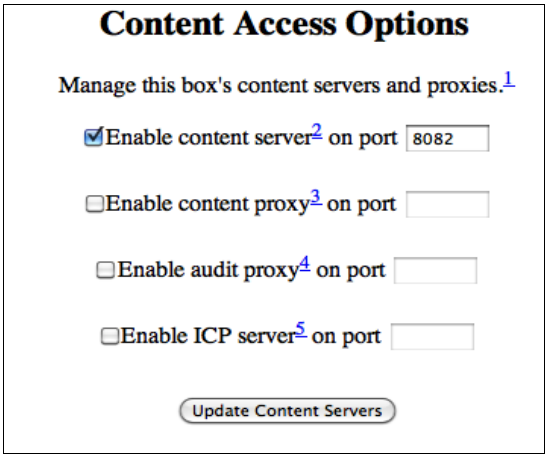
LOCKSS (Lots of Copies Keep Stuff Safe) and Ex Libris have made content preserved in LOCKSS Boxes available through online public access catalogs (OPACs) that use the Ex Libris SFX link resolver. This guide describes the steps required to customize your LOCKSS Box and SFX instance.

Customizing Your LOCKSS Box

A LOCKSS Box has an optional Content Server that makes preserved content available from a web browser through OpenURL queries. The SFX LOCKSS target uses OpenURL queries to request content from your local LOCKSS Box. The Content Server is not enabled by default. To make content in your LOCKSS Box available through SFX, use its configuration screens to enable and configure this feature.

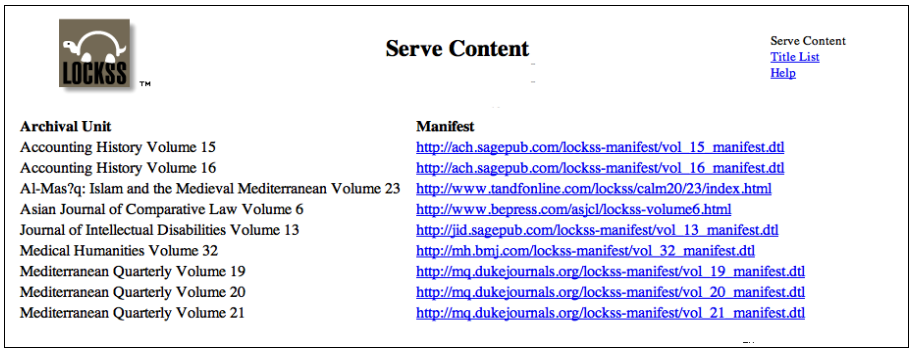
Enabling the Content Server

From the main administration screen of your LOCKSS Box, select **Content Access Options**. Then select **Content Server Options**. This will take you to a screen for managing the LOCKSS Box's content servers and proxies. Select **Enable content server on port** and enter the port number to use for serving preserved content from a web browser. Port numbers below 1024 require the application to run as root and are therefore not available. It is recommended that you choose a port that is close to 8081, the default port used by the administration GUI of of your LOCKSS Box. Port 8082 is used in this example.



Finally, select **Update Content Servers** to finalize this change. The operating system on your LOCKSS Box host may require opening non-standard ports to enable outside access. Consult with the person who administers the host where your LOCKSS Box is running to determine which standard ports are available.

To test your Content Server configuration, use your browser to contact the Content Server of your LOCKSS Box which you set up above. By default, the Content Server displays a list of Archival Units (AUs) that are configured. In this example, if your LOCKSS Box administration GUI is accessed at http://lockss.xyz.edu:8081/, your Content Server is at http://lockss.xyz.edu:8082/ServeContent.



Serve Content

Serve Content
Title List
Help

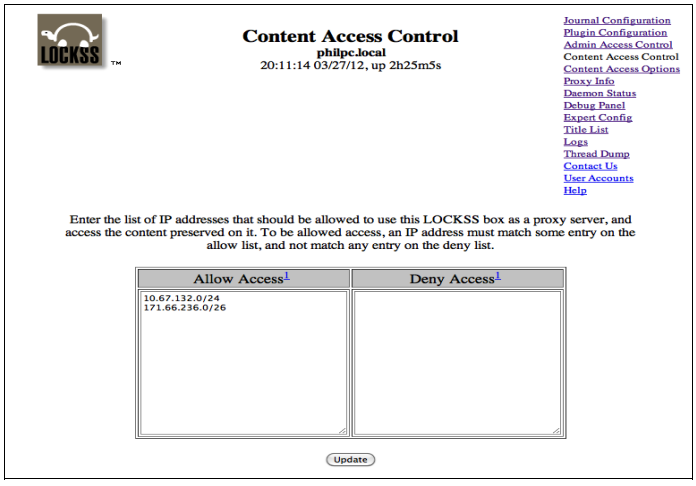
Archival Unit	Manifest
Accounting History Volume 15	http://ach.sagepub.com/lockss-manifest/vol_15_manifest.dtl
Accounting History Volume 16	http://ach.sagepub.com/lockss-manifest/vol_16_manifest.dtl
Al-Mas?q: Islam and the Medieval Mediterranean Volume 23	http://www.tandfonline.com/lockss/calm20/23/index.html
Asian Journal of Comparative Law Volume 6	http://www.bepress.com/asjcl/lockss-volume6.html
Journal of Intellectual Disabilities Volume 13	http://jid.sagepub.com/lockss-manifest/vol_13_manifest.dtl
Medical Humanities Volume 32	http://mh.bmj.com/lockss-manifest/vol_32_manifest.dtl
Mediterranean Quarterly Volume 19	http://mq.dukejournals.org/lockss-manifest/vol_19_manifest.dtl
Mediterranean Quarterly Volume 20	http://mq.dukejournals.org/lockss-manifest/vol_20_manifest.dtl
Mediterranean Quarterly Volume 21	http://mq.dukejournals.org/lockss-manifest/vol_21_manifest.dtl

You must ensure that the packet filters (i.e. firewall) configured on your LOCKSS Box allow access to the port you chose for your Content Server. The default packet filter is 8082. Enabling the port for access outside of the network requires administrative privileges, and depends on the operating system running your LOCKSS Box. Ask the systems administrator to do this.

Configuring Content Server Access Control

Now that you have enabled the Content Server, you'll need to give your user community access to preserved content. To do this, return to the main administration page of your LOCKSS Box and select **Content Access Control**. You will see two lists side-by-side. The "Allow Access" list specifies which IP addresses or ranges can access preserved content through the Content Server. The **Deny Access** list allows you to deny access to specific IP addresses or ranges, typically ones within a wider range specified by the **Allow Access** list.

The default **Allow Access** list includes two address ranges: the initial network address that was specified when your LOCKSS Box was installed, and those by the LOCKSS organization to provide technical support.



Content Access Control
philpc.local
20:11:14 03/27/12, up 2h25m5s

Journal Configuration
Plugin Configuration
Admin Access Control
Content Access Control
Content Access Options
Proxy Info
Daemon Status
Debug Panel
Expert Config
Title List
Logs
Thread Dump
Contact Us
User Accounts
Help

Enter the list of IP addresses that should be allowed to use this LOCKSS box as a proxy server, and access the content preserved on it. To be allowed access, an IP address must match some entry on the allow list, and not match any entry on the deny list.

Allow Access[1]	Deny Access[1]
10.67.132.0/24 171.66.236.0/26	

Update

Consult your contracts with publishers whose content is preserved on your LOCKSS Box to determine what IP address ranges have access to them. You should configure the **Allow Access** list to include the same IP address ranges that you supplied to publishers.

Note: The simple configuration that is supported by your LOCKSS Box assumes that all publishers support the same IP address ranges.

For example, XYZZY University provides the following IP address ranges to its e-pub vendors:

IP Address	CIDR Prefix	Netmask	Use
10.12.0.0	/16	255.255.0.0	Dorms
10.64.0.0	/16	255.255.0.0	Main Campus
10.65.0.0	/16	255.255.0.0	Medical Center

Your institution's contracts with publishers may exclude certain campus functions. Only those address ranges that are allowed should be included in the **Allow Access** list. Here is what XYZZY University would add to its list:

```
10.12.0.0/16
10.64.0.0/16
10.65.0.0/16
```

Once you are done configuring the access control for your LOCKSS Box, select **Update** to finalize your configuration. Test your configuration by accessing the Content Server from various allowed and denied hosts within your institution.

Using Apache as a Front End to your LOCKSS Content Server

The configuration just described gives end users direct access to your LOCKSS Box's Content Server through the port you specified. Some IT organizations limit the use of “non-standard” ports for web services because it requires opening additional ports in their institution's firewalls. These organizations may also prefer to administer access using the same web server access controls that they use throughout their institution. If this does not apply to your institution, you can skip this section.

One way to accomplish this is to direct users to a web server such as Apache that runs on the standard port 80, and configure the Apache server to act as a proxy. It relays requests to the LOCKSS Box Content Server and returns responses from the LOCKSS Box to the user. This allows the presence of the LOCKSS Box to be hidden, and enables the IT staff to use existing access control configurations on the Apache server to limit access to preserved content, including user authentication using HTTPS and SSL certificates.

Load balancing could also be done across several LOCKSS Boxes in a similar way. The Apache instance can be one that is shared by many applications within your institution, or it can be a custom instance that is installed on the same host as the LOCKSS Box. The simplest configuration uses the standard 'mod_proxy' Apache module to forward content server traffic to and from the LOCKSS Box. Here is a typical 'mod_proxy' configuration for an Apache server virtual host at port 80 on the same machine as the LOCKSS Box:

```
<VirtualHost *:80>
ServerAdmin lockss-admin@xyzzy.edu
DocumentRoot /var/www/html
ServerName lockss.xyzzy.edu
<IfModule mod_proxy.c>
ProxyRequests Off
ProxyVia On
<Proxy lockss.xyzzy.edu/*>
AddDefaultCharset off
Order deny,allow
Allow from all
</Proxy>
ProxyPassMatch ^/((ServeContent|images).*)$
http://localhost:8082/$1
</IfModule>
ErrorLog logs/error_log
CustomLog logs/access_log common
</VirtualHost>
```

If you use this technique, you should omit adding IP addresses to the **Content Access Control** of your LOCKSS Boxes because it will only be accessed locally by the Apache server. If the Apache server is on a different subnet, you should add only the IP address of that host to the Content Access Control of your LOCKSS Box. You should also add the following special configuration parameter to **Expert Config** screen of your LOCKSS Box.

```
org.lockss.serveContent.absoluteLinks=false
```

This causes Content Server to rewrite links in pages it serves relative to the LOCKSS Box address, which simplifies the 'mod_proxy' configuration.

Customizing your SFX Instance

Once you have configured your LOCKSS Box, the next step is to make its contents available through SFX. In SFX, you will need to configure the LOCKSS target and target service, and activate portfolios that correspond to content that is available through your LOCKSS Box.

Configuring the LOCKSS Target and Target Service in SFX

First, locate the LOCKSS target in SFX and activate it. You may want to add a General Note or an Authentication Note to indicate the conditions under which content is available. For XYZZY University, the message might be, "On campus access only."

Next, activate the getFullTxt service of the LOCKSS target. At the target service level, you will need to make a variety of adjustments. Be sure to edit the L/P parameters to match the details of your LOCKSS Box. In the example shown earlier, where the LOCKSS Box is accessed directly, the $$HOST parameter should be lockss.xyzzy.edu and the $$PORT parameter should be 8082.If you are using an Apache web server on the same machine, with a proxy at the standard web port 80, the $$HOST parameter would still be lockss.xyzzy.edu but the $$PORT parameter would be 80.

Ex Libris updates the LOCKSS target in its KnowledgeBase monthly, based on information they receive from LOCKSS about titles and volume ranges that are available for preservation. We recommend that you configure the LOCKSS target for automatic updates by activating the "AutoUpdate" checkbox in the LOCKSS target service.

At present, LOCKSS only provides the getFullTxt service for electronic resources, so that will be the only target service you can activate. Services such as getAbstract and getTOC will be available in the future.

Activating Object Portfolios

Now you are ready to activate the appropriate LOCKSS portfolios in your SFX instance. The following paragraphs describe four options for activating the portfolios. The best choice for your installation will depend on a number of factors.

Activate all available portfolios.

At the target service level, select "AutoActive" and at the portfolio level, select the "Activate All" button. This ensures that new LOCKSS portfolios that arrive from Ex Libris will be automatically activated. The LOCKSS Box Content Server will redirect user requests to the publisher if the content they request is not preserved in your local LOCKSS Box.

This option works well if your SFX instance is **not** used to determine the content of your OPAC. If your OPAC records are the holdings authority of your library's online content, a LOCKSS portfolio would only be offered to users where an OPAC record already exists for the title.

This option **does not** work well if your library uses your active SFX portfolios as your holdings authority.

For example, if you use the Ex Libris MARCIt! service, your active SFX portfolios are used to generate records that could end up in your OPAC. If you activate all the portfolios in the LOCKSS target, then all of them would end up as OPAC records, including ones that you do not have have access to at your institution. A similar thing would happen for other systems that are integrated with SFX and assume that all your active portfolios are resources actually held by your library.

Activate only configured portfolios.

This option uses the list of AUs that you have configured in your LOCKS box to activate LOCKSS portfolios, which allows you to activate LOCKSS portfolios in your SFX instance at the same time you configure the corresponding AUs in your LOCKSS Box. Use the List Holdings feature of your LOCKSS Box to export a list title_ids (ISSNs or EISSNs), and coverage notes using the **Configured** and **Titles** options. Using this exported list, you can activate corresponding portfolios in SFX and either verify or create thresholds for those portfolios.

This option works well if you manually configure your LOCKSS box or use the Subscription Manager feature starting in LOCKSS version 1.62 to automatically configure your LOCKSS Box for only those AUs that your library can access from the publisher. It takes time for your LOCKSS Box to crawl configured content, and users are automatically redirected to the publisher while content is being collected. However, the LOCKSS Box automatically starts serving the preserved content once the collection is complete.

This option **does not** work well if your library simply configures every AU available for preservation and relies on crawls succeeding or failing to determine what your LOCKSS Box can actually access. It would be far easier in that case case to choose option 1, and avoid having to keep manually updating the assets in your SFX instance from the list of configured AUs.

Activate only collected portfolios.

This option uses the list of AUs that have actually been collected by your LOCKSS Box to activate LOCKSS portfolios. Use the List Holdings feature of your LOCKSS Box with the **Collected** and **Titles** options to export a list if title_ids (ISSNs or EISSNs), and coverage notes for only those AUs that have been successfully ingested on your LOCKSS Box.

This option works well if you normally configure every AU available for preservation on your LOCKSS Box and rely on crawls succeeding or failing to determine what your LOCKSS Box can actually access. Only those portfolios that correspond to AUs that have actually been collected by your LOCKSS Box will be activated in your SFX instance.

The disadvantage is that you can only activate portfolios as corresponding AUs are collected by your LOCKSS Box, rather than when they are configured. It takes much more

monitoring to determine which AUs have become ready since the last time you activated LOCKSS portfolios from the "collected" title list.

Activate only discontinued portfolios.

This option only enables LOCKSS portfolios for titles where your institution has either discontinued its subscription to a publication or the publication is no longer published, and the publication is not available from the publisher. You use the **List Holdings** feature of your LOCKSS Box with the **Collected** and **Titles** options to export a list ISSNs or EISSNs, and their coverage notes for only those AUs that have been successfully ingested on your LOCKSS Box. Then you activate only those titles to which you used to subscribe but are no longer available from the publishers.

This option works well if you want to limit the number of targets that are offered to your users, and prefer to show only the publisher target if it is available.

This option **does not** work when the content becomes temporarily unavailable from the publisher, because it requires you to actively manage your targets (i.e. temporarily disabling the publisher target and enabling the LOCKSS target during the outage, and then reversing the process once the publisher content is available once again). Your LOCKSS Box will automatically manage this process if you always offer your users the LOCKSS target.

Using DataLoader to Activate Object Portfolios

The new SFX DataLoader option now enables you to use the SFX DataLoader to automate the process of activating object portfolios and creating thresholds using the **Configured** or **Collected Title List** listing produced by your LOCKSS Box. DataLoader requires a text file with tab-delimited fields. Input should consist of a single line for each portfolio. The line must include an identifier for the portfolio. For example, using DataLoader, you could load a text file into SFX, where the each line includes an ISSN. Besides the identifier field, you may have optional fields that represent the activation status and threshold of each portfolio.

Beginning in the 1.60 version of LOCKSS, the Title List provides an easy option for exporting to a format that can be used with the SFX DataLoader utility. For more information about activating portfolios using DataLoader, see the SFX "General User's Guide."

First, configure the Title List in the LOCKSS Box to produce a report by either selecting **Configured** or **Collected** content, and the **SFX Data Loader** format.

The **SFX DataLoader** format outputs coverage ranges in the "coverage_notes" field of the report in a format that are used as thresholds by the DataLoader. For example, let's say a row in your custom report has the values "2000 1 2010 30" which means the title started in 2000 with volume 1 and ended in 2010 with volume 30. This range will be formatted in the "coverage_notes" field as

```
$obj->parsedDate(">=",2000,1,undef) && $obj-
>parsedDate("<=",2010,30,undef)
```

If there are multiple ranges, they will be separated by two vertical bars:

```
$obj->parsedDate(">=",2004,1,undef) && $obj-
>parsedDate("<=",2006,5,undef) ||

$obj->parsedDate(">=",2006,7,undef) && $obj-
>parsedDate("<=",2009,10,undef)
```

Select the **List Titles** button to generate the report. You can now load the new file into SFX. In DataLoader, select the LOCKSS target. Choose either ISSN or ISBN for column 1, depending on the type of data you are loading.

Choose ACTIVATION_STATUS for column 2, and THRESHOLD for column 3. From the check-boxes at the bottom of the screen, select **Update Portfolios**. (Do not select **Add Portfolios** because that may create unwanted local data.) After the loading process, check the resulting error files to see if any portfolios failed to update. You may discover that some portfolios need additional editing to ensure SFX linking.

Finally, in the KnowledgeBase, click the Test buttons of some of your newly updated LOCKSS portfolios to check if you reach LOCKSS.

Troubleshooting

If you cannot access an article from SFX, here is a sequence of steps that you can take to diagnose the problem.

1. Verify that the LOCKSS target is properly configured in your SFX instance.
2. Temporarily set the following configuration parameter in your LOCKSS Box **Expert Config** screen. This causes the OpenURL resolver to output detailed information about how it handles queries.org.lockss.logOpenUrlResolver.level=debug3
3. Send an OpenURL query to your LOCKSS Box Content Server for an article that is preserved. Here is a sample OpenURL query: http://lockss.xyzzy.org:8082/ServeContent?issn=1085-4908&volume=25&issue=1&spage=7
4. Check your LOCKSS log file by selecting the Logs link in the list of actions on the right of your LOCKSS Box administrative GUI, and the the 'daemon' file from the log directory list. The most recent lines should display OpenUrlResolver log information about the query.

If you cannot determine the cause of the problem, contact LOCKSS support at support@lockss.org.