Keeping your Archive Safe (and on TRAC) with SafeArchive and LOCKSS

Micah Altman <micah_altman@alumni.brown.edu>
Jonathan Crabtree <jonathan_crabtree@unc.edu>
Thu-Mai Christian <tlchristian@unc.edu>

Location 311

June 5, 2012 | 9:30am – 12:30pm

Overview

This ½-day workshop focuses on how to protect your archival content, and how to formalize, document and audit storage policies. The workshop is appropriate for curators of content who wish to replicate their content and/or document storage policies for compliance with TRAC, Data Seal of Approval, and other archival standards. This provides hands-on practice with and example configurations for LOCKSS installation and configuration; SafeArchive installation and configuration, generating and interpreting policy reports, and TRAC documentation. At the end of the workshop the student will be able to install the SafeArchive system; use it to replicate archival content exposed through DVN, OAI-PMH or the web; and produce formal audits and reports to determine compliance with archival storage policies.

Requirements

- Laptop computer with the latest version of Firefox installed
- Amazon Web Services account (please obtain an AWS account *prior* to the workshop event)

Agenda

Introductions	9:30 – 9:45
Overview and Tour of the SafeArchive System	9:45 – 10:15
Hands On: Running the SafeArchive System using Amazon Web Services	10:15 – 11:00
Break	11:00 – 11:10
Hands On: Setting up the SafeArchive System	11:10 – 11:40
Hands On: Using the SafeArchive System	11:40 – 12:30

Running the SafeArchive System Using Amazon Web Services

The SafeArchive System (SAAS) can easily be run using Amazon Web Services. While SAAS is free-to-use open source software, Amazon Web Services (AWS) charges a fee for web hosting and data storage services. For this workshop, you can expect to incur a minimal cost of less than $5. *To avoid any additional fees, you must ensure that you terminate the service at the conclusion of the workshop*.

Amazon's Getting Starting guide at http://docs.amazonwebservices.com/AWSEC2/latest/GettingStartedGuide may also be helpful as you set up the SAAS system using AWS.

Launching the SafeArchive Amazon Machine Instance

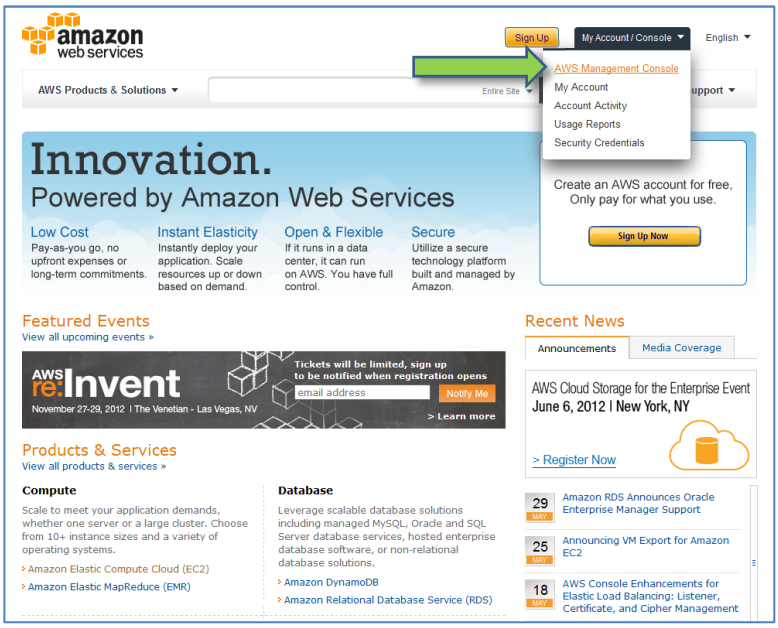
The SafeArchive Installation begins in the AWS Management Console within Amazon Web Services. The instructions below provide guidance on launching and running the SafeArchive Amazon Machine Instance (AMI).

---

1. Launch a supported web browser such as Firefox.

---

2. Navigate to the Amazon Web Services homepage: http://aws.amazon.com. Sign in to the AWS Management console by clicking on the **My Account/Console** button and selecting **AWS Management Console**.

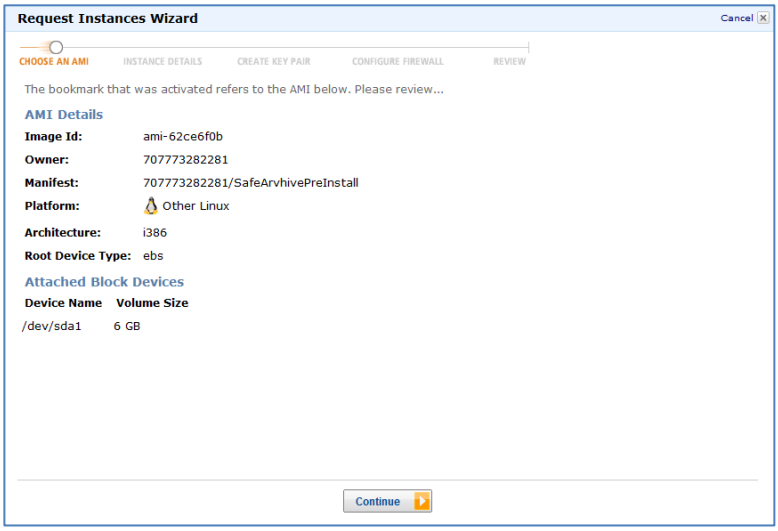


---

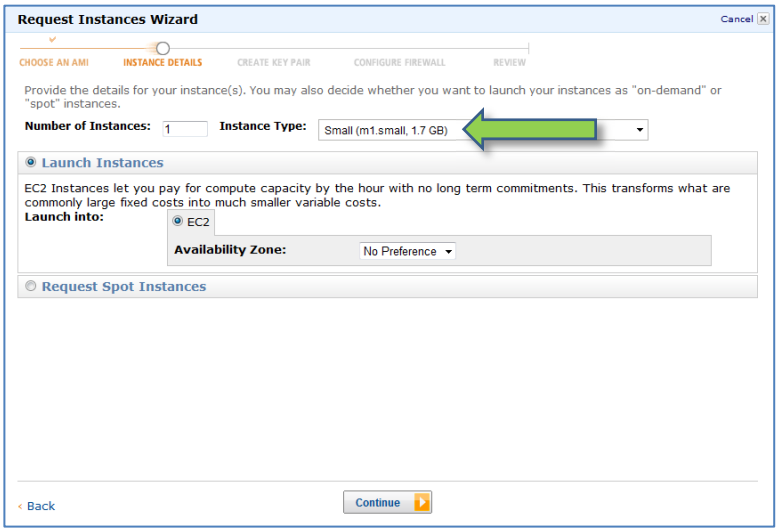
3. Enter your e-mail address and Amazon password. If you do not have an Amazon account, you can create one for free by selecting "I am a new user."

If you have not yet signed up for AWS, a screen will appear that will instruct you through the process of signing up for AWS.

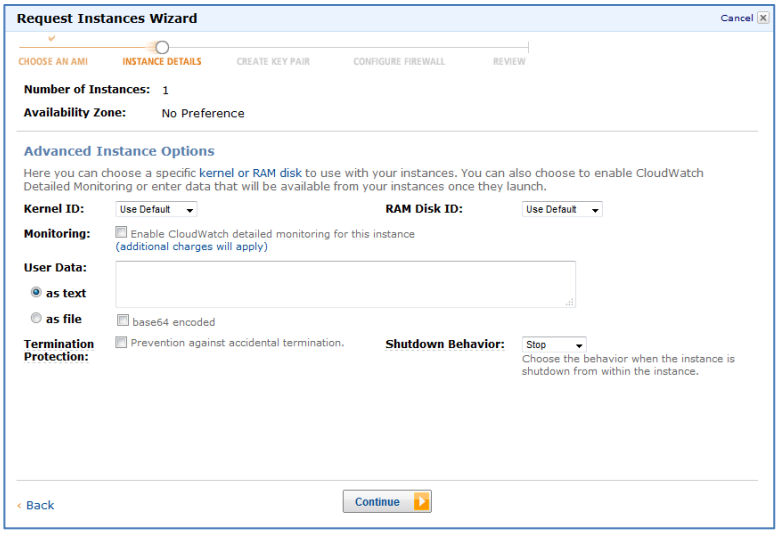
4. Follow the SAAS AMI link <https://console.aws.amazon.com/ec2/home?region=us-east-1#launchAmi=ami-e67ddc8f > to request the SafeArchive Installation AMI. When the Request Instances Wizard dialogue box appears, click on the **Continue** button.



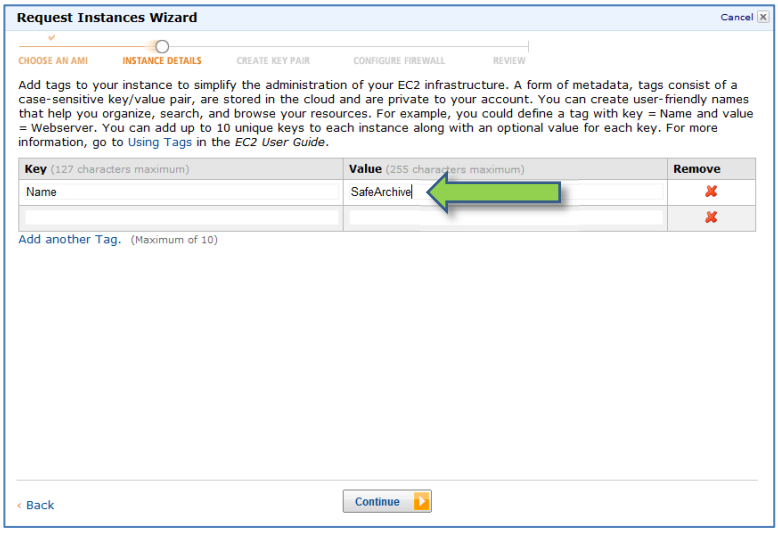
5. In the Instance Details step of the Request Instances Wizard, change the Instance type to "Small (m1.small, 1.7GB)." Click on the **Continue** button.



6. On the next screen, keep the default settings. Click on the **Continue** button.

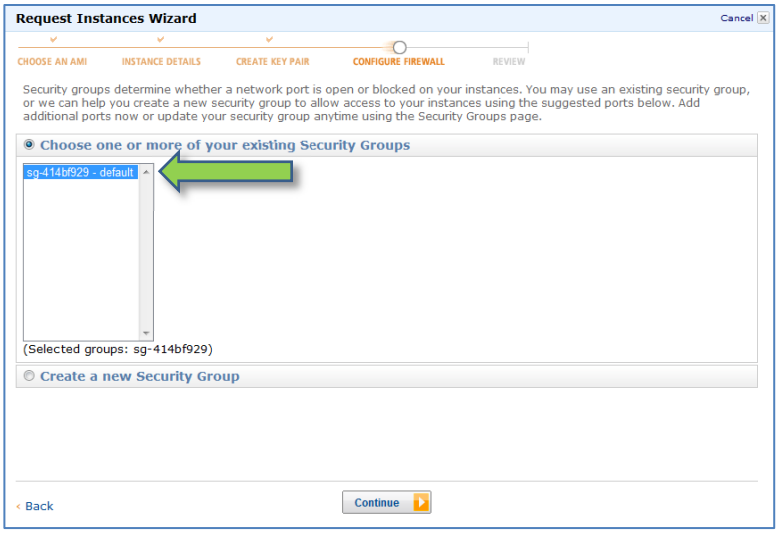


7. Enter a Value for the “Name” Key, which will be used to identify your SAAS instance. Click on the **Continue** button.

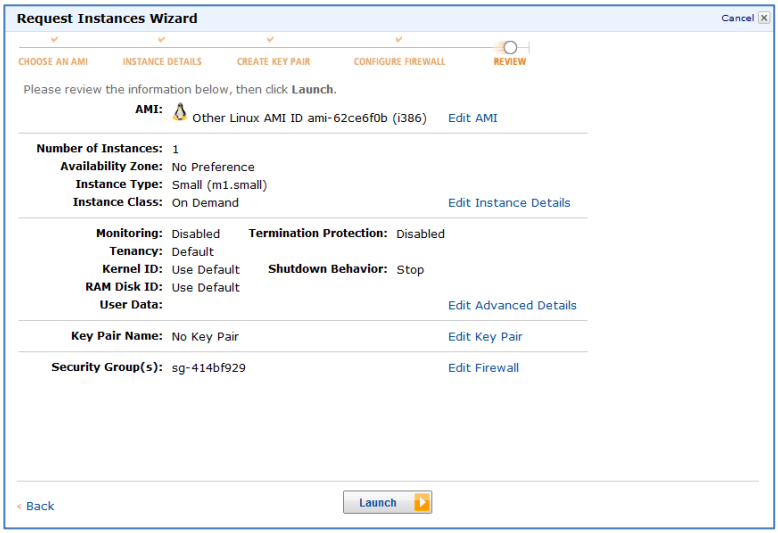


8. The SAAS Installer has been programmed to operate without the need for a new Key Pair. Select “Proceed without a Key Pair.” Click on the **Continue** button.

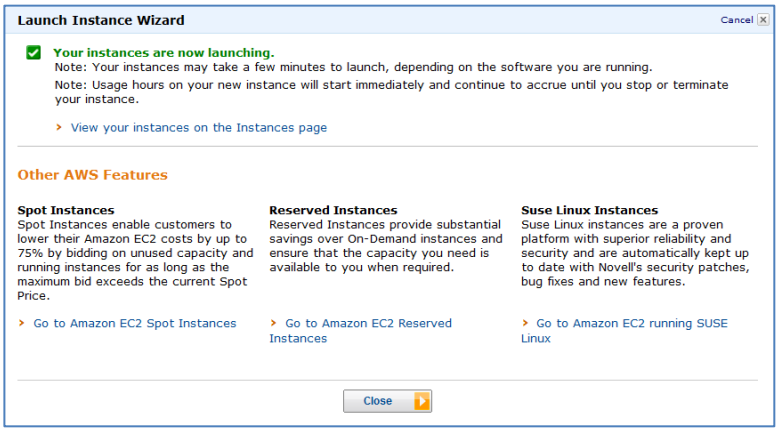
9. In the Configure Firewall step of the Request Instances Wizard, select the default Security Group. Click on the **Continue** button.



10. Click on the **Launch** button to accept your settings and proceed.



11. The SAAS Instance request process is now complete. **Close** the Launch Instance Wizard dialogue box.



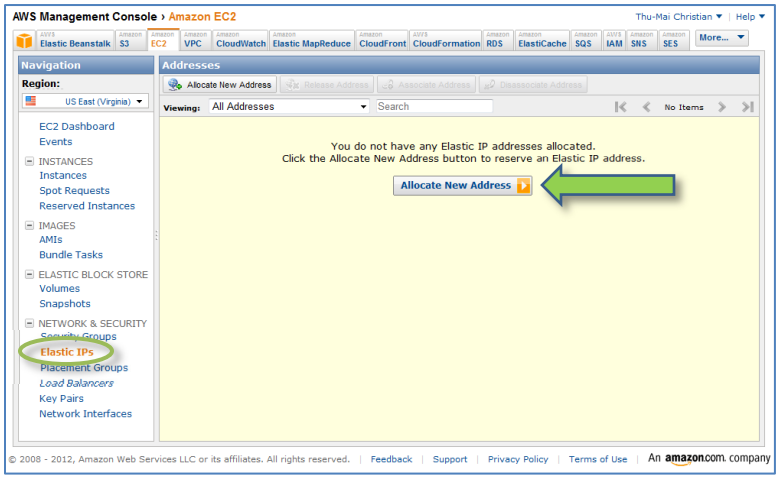
Configuring and Installing the SafeArchive AMI

Once the SafeArchive AMI is running, it will need to be associated with an IP address and TCP ports must be specified prior to running the SafeArchive Installer. The instructions below provide guidance on completing these required tasks.

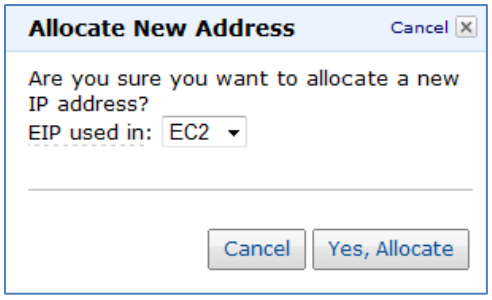
1. Creating an Elastic IP
2. Setting up Security Groups
3. Running the SAAS Installer

---

1. Your SAAS installation requires that an IP address be associated with the instance. To generate an IP address, click on the **Elastic IPs** link in the Navigation panel on the left-hand side of the screen. Click on the **Allocate New Addresses** button at the top of the Addresses panel.

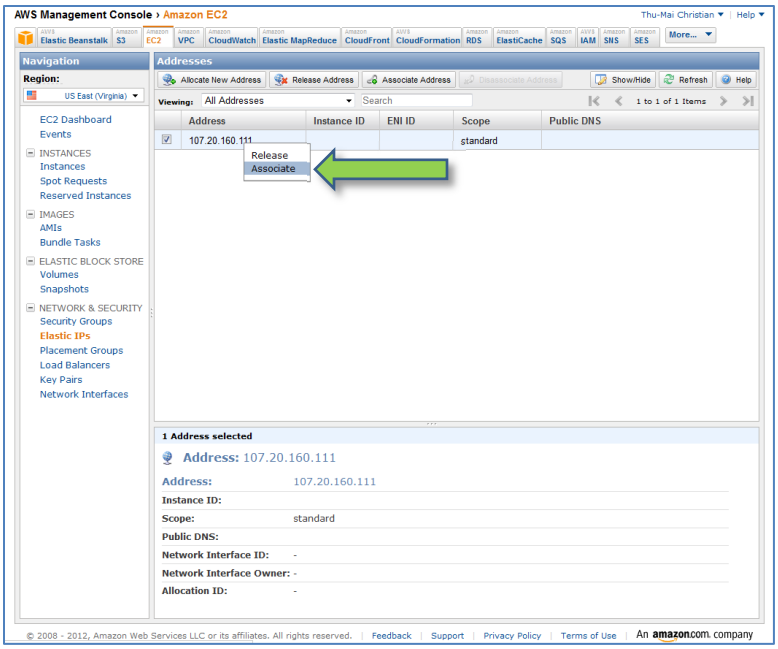


A dialogue box will appear to confirm the IP allocation action. Click on the **Yes, Allocate** button.

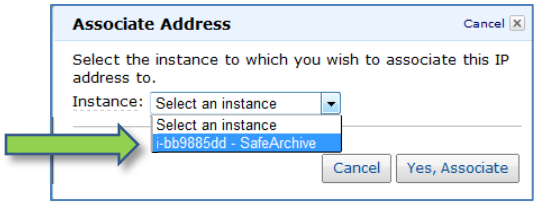


Associate your new elastic IP by right-clicking on the IP address you just generated. Select **Associate**.

Record your new Elastic IP Address here:



A dialogue box will appear, which allows you to select the instance to which the IP address will be associated. Click on the SafeArchive AMI. Click on the **Yes, Associate** button.



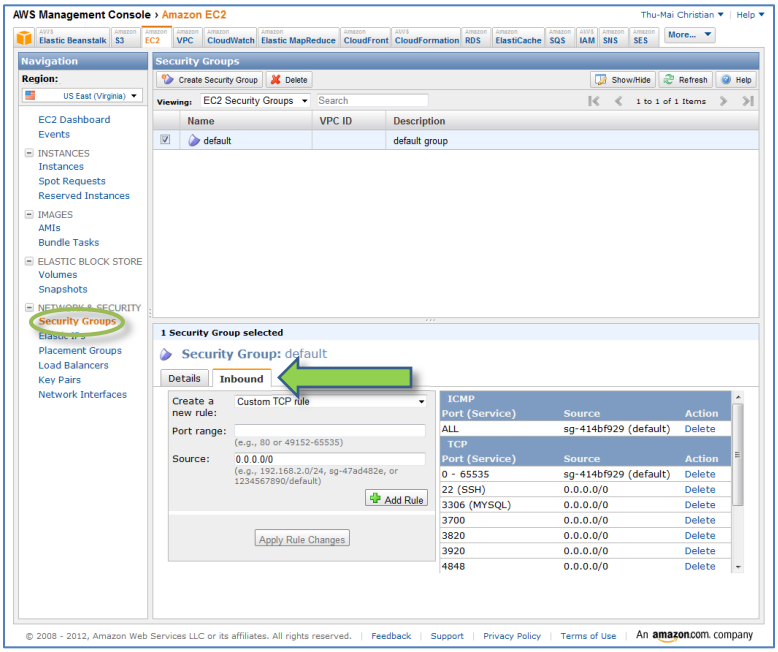
---

2. Running SAAS requires several TCP ports to be open. To specify these ports, click on the **Security Groups** link in the Navigation panel on the left-hand side of the screen. Check the box to the left of your designated security group (the "default" security group will be used for the workshop). In the panel below, click on the **Inbound** tab. In the **Port range** field, enter the port number. Click on the **+Add Rule** button. Complete this process for each of the 8 ports listed in the box below.

22	3306	3700	3820
3920	4848	8080	8686

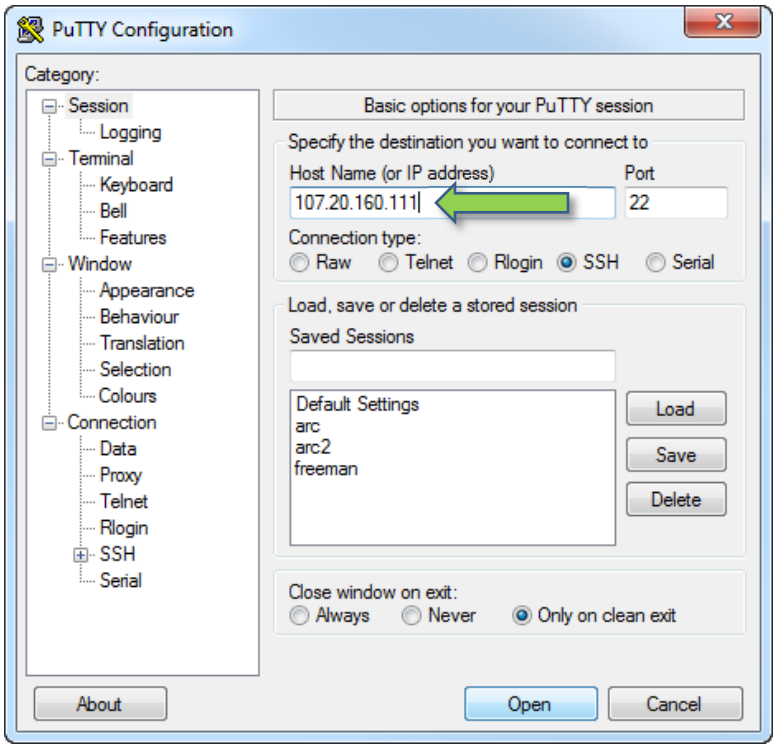
TCP ports for SAAS installation

When you have addded all of the ports, click on the **Apply Rule Changes** button.



3. To run the SAAS installer, open PuTTY.exe.

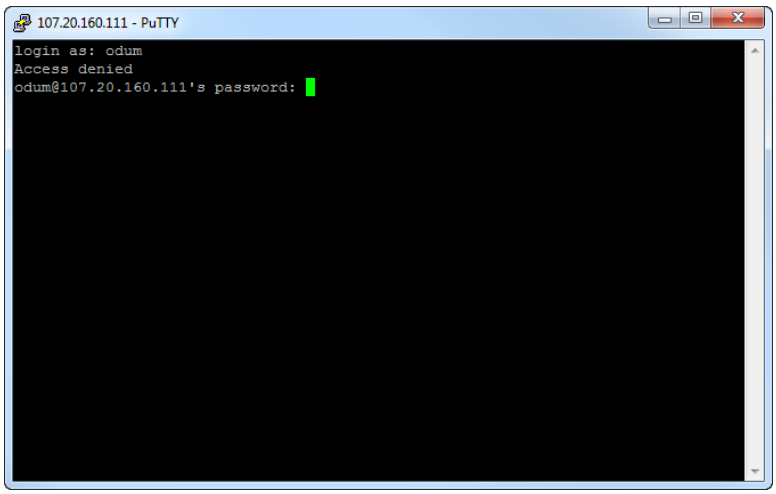
In PuTTY, enter the elastic IP address you allocated in the AWS Management Console in the Host Name (or IP address) field. Click on the **Open** button.



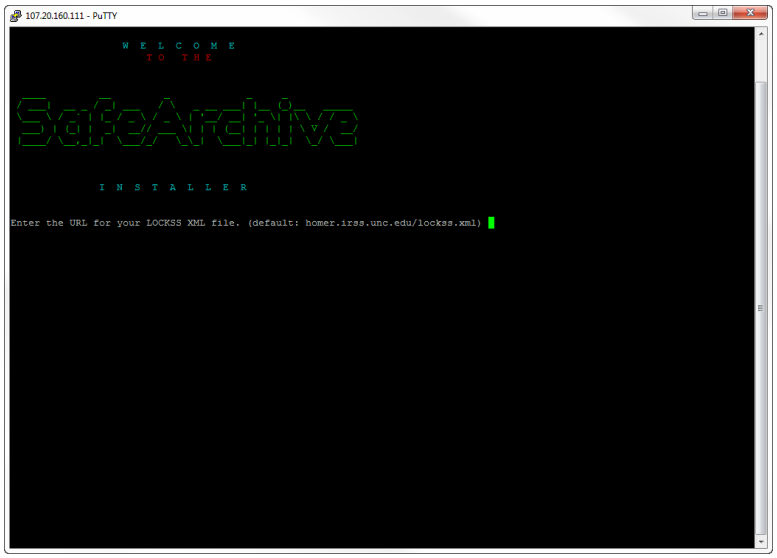
PuTTY will launch a shell script that will prompt you for a login and password. Enter the following information:

login: odum
password: odum

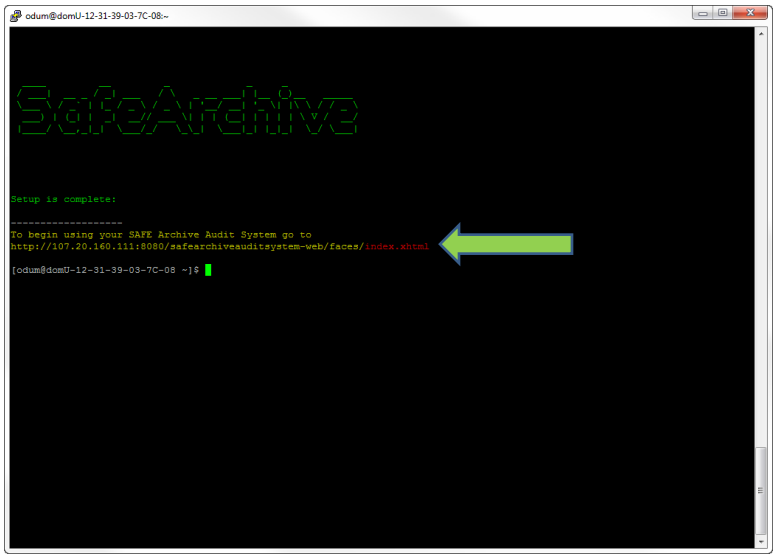
SAAS Installer login & password



The SafeArchive installer will prompt you for a series of inputs to configure the SafeArchive System. Default inputs are provided. To accept the default, press the enter key at the prompt.



Once the system has loaded all necessary applications and configurations, the installer will indicate that Setup is complete. A URL is provided, which links to your SafeArchive instance. Enter this URL into a Firefox browser address bar to launch the SafeArchive System.



Setting Up the SafeArchive System

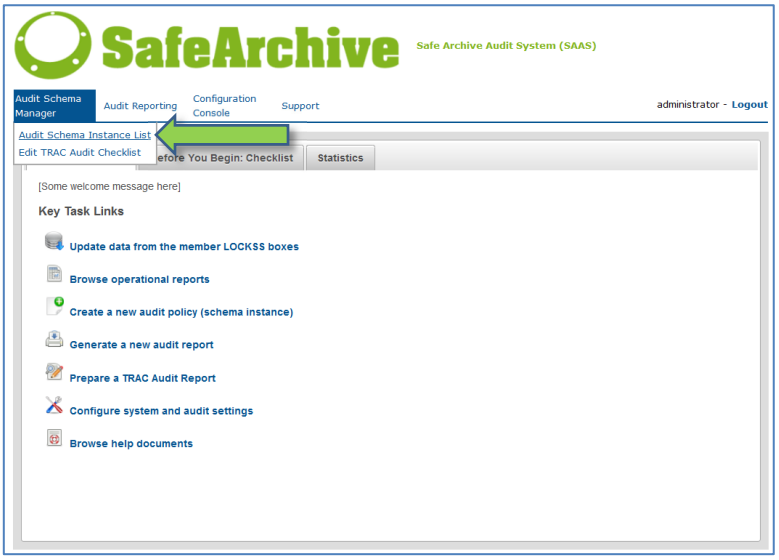
This section of the workshop will enable you to complete the following tasks:

1. Navigating to the Audit Schema Instance List page
2. Adding your owner-institution name
3. Specifying subject terms
4. Refreshing the network data status tables
5. Adding an archival unit

Prior to using the SAAS, several steps must be completed. These steps include establishing your owner-institution, providing a subject list, adding hosts and archival units, editing host and archival unit information, and refreshing the network status data tables. The instructions below provide guidance on accomplishing these required tasks.

---

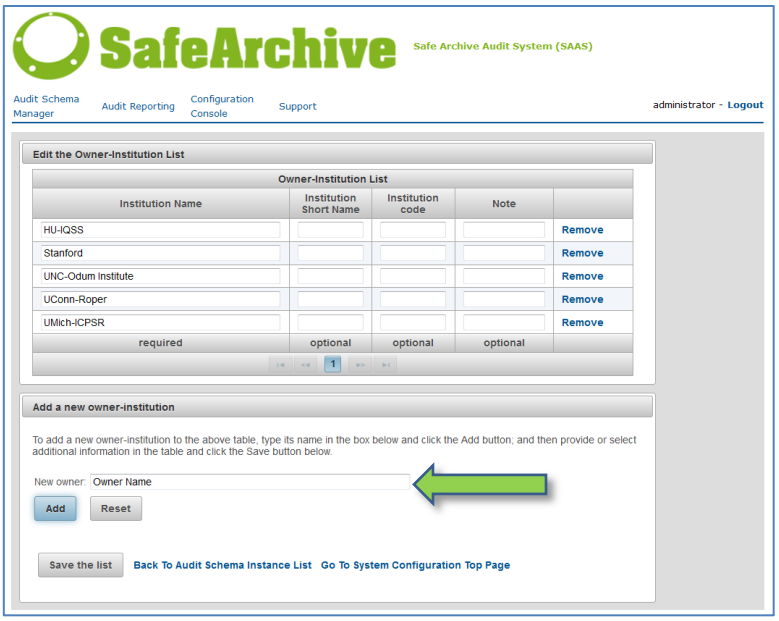
1. Navigate to the Audit Schema Instance List page by selecting the **Audit Schema Instance List** link in the **Audit Schema Manager** menu at the top of the screen.



---

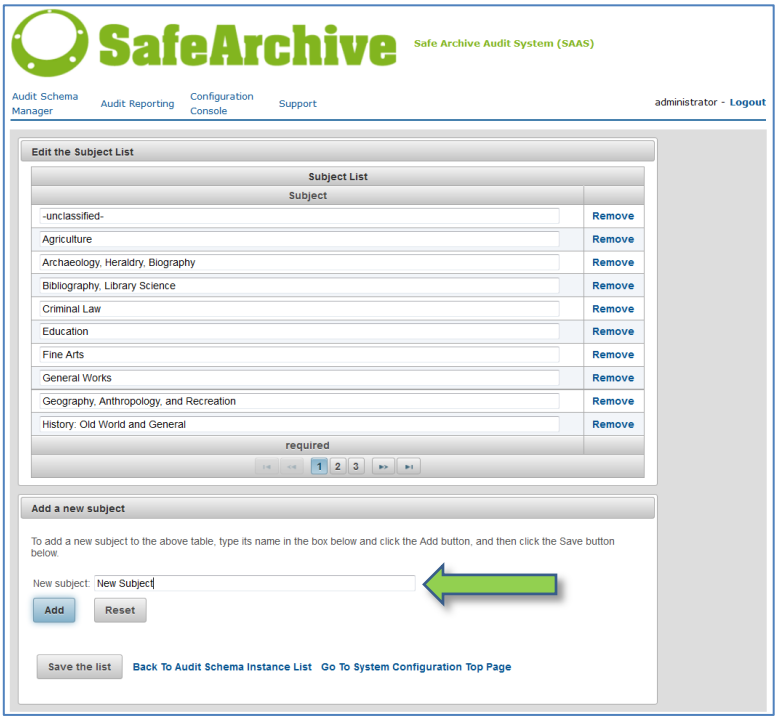
2. Add your owner-institution name to the system by clicking on the **Edit the Owner-Institution Option List now** link.

In the **New owner** field, enter your owner-institution name. Click on the **Add** button. Once your institution appears in the table, enter text in the optional fields next to your Institution Name. **Click on the Save** the list button. Return to the Audit Schema Instance List page.

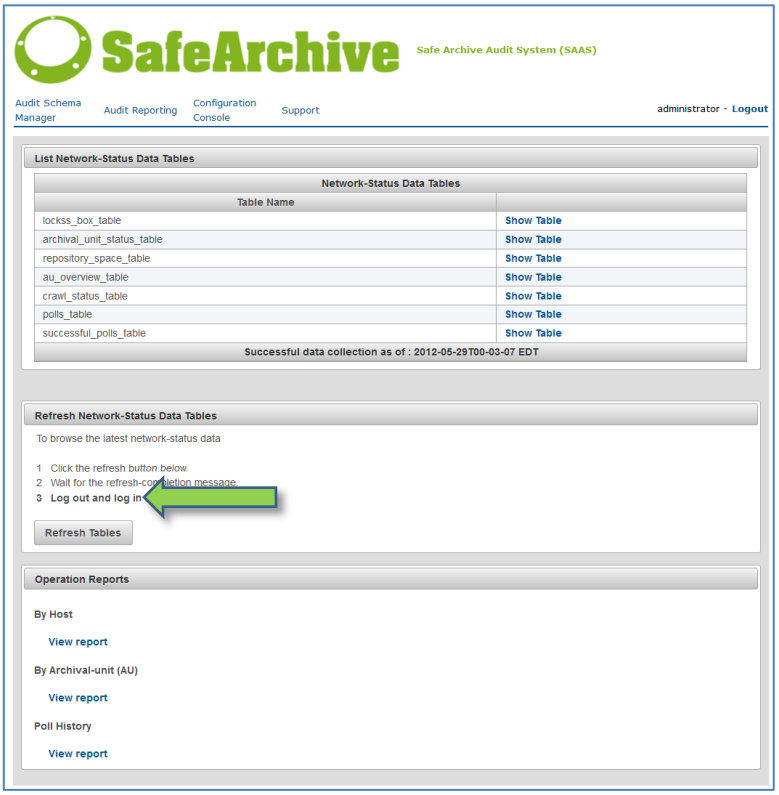


---

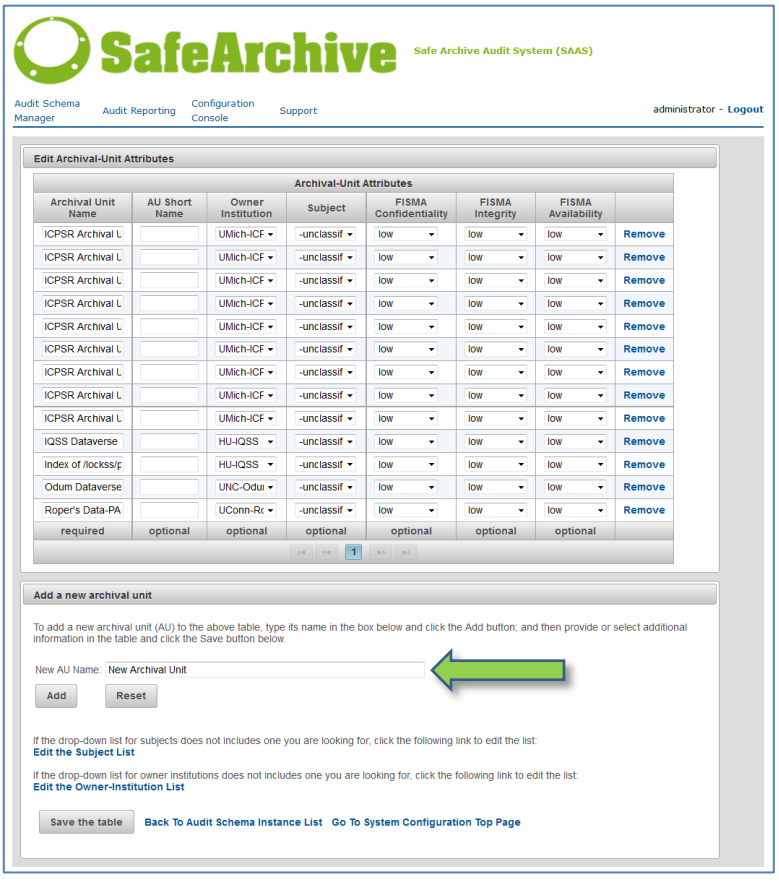
3. The Subject Option List allows you to specify subject terms used to describe archival units. To edit and/or add subject terms, click on the **Edit the Subject Option List** now link on the Audit Schema Instance List page. Add a new subject by entering a term in the **New subject** field. Your term will appear in the Subject List table after you click on the **Add** button. Click on **the Save the list** button to save your updated Subject List table. Return to the Audit Schema Instance List page.



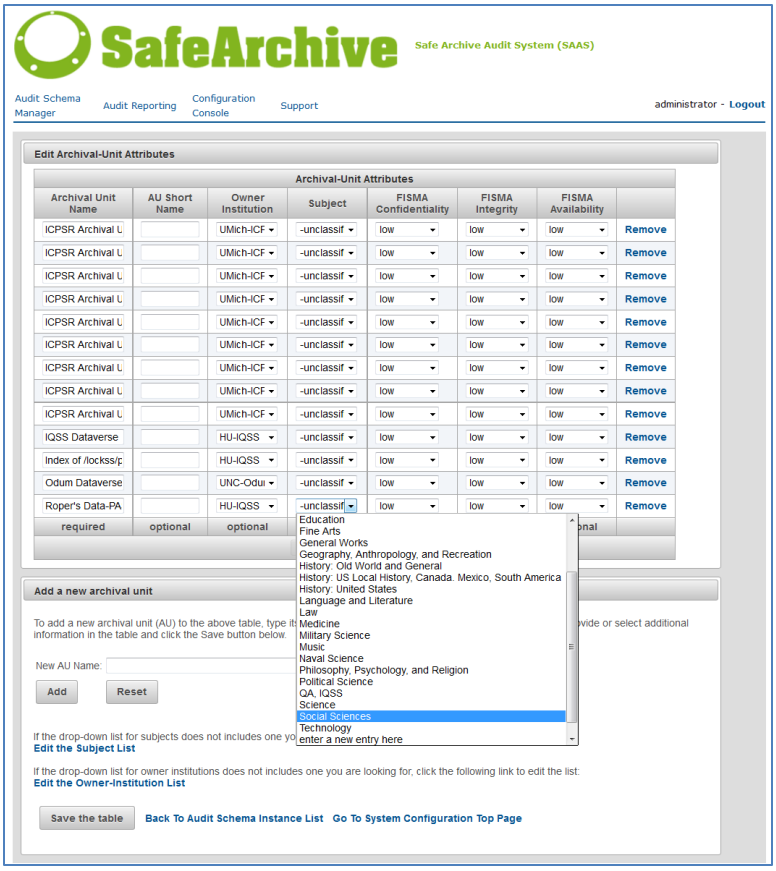
4. Refresh the network data tables. From the Audit Reporting menu option at the top of the screen, select **Detailed Network Status Data**. Click on the **Refresh Tables** button. Log out of the SAAS and log in again.



5. Add a new archival unit by clicking on the **Edit Archival-unit attributes now** link on the Audit Schema Instance List page. Enter the name of the archival unit in the **New AU Name** field. Click on the **Add** button.



Once the new archival unit appears in the table, you can enter an AU Short Name by entering text into the text field next to your archival unit name. You may also specify the Owner Institution, Subject, and levels of FISMA Confidentiality, Integrity, and Availability by selecting from the option in the drop down boxes next to your archival unit name. Click on the **Save the table** button to save your selections.



Creating Preservation Policies

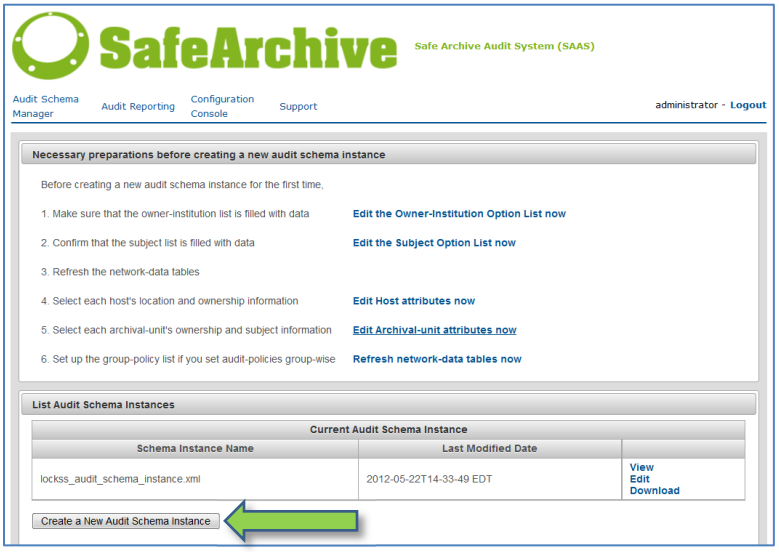
This section of the workshop will enable you to complete the following tasks:

1. **Establishing Network Audit Policy**
2. **Completing the TRAC Audit Checklist**

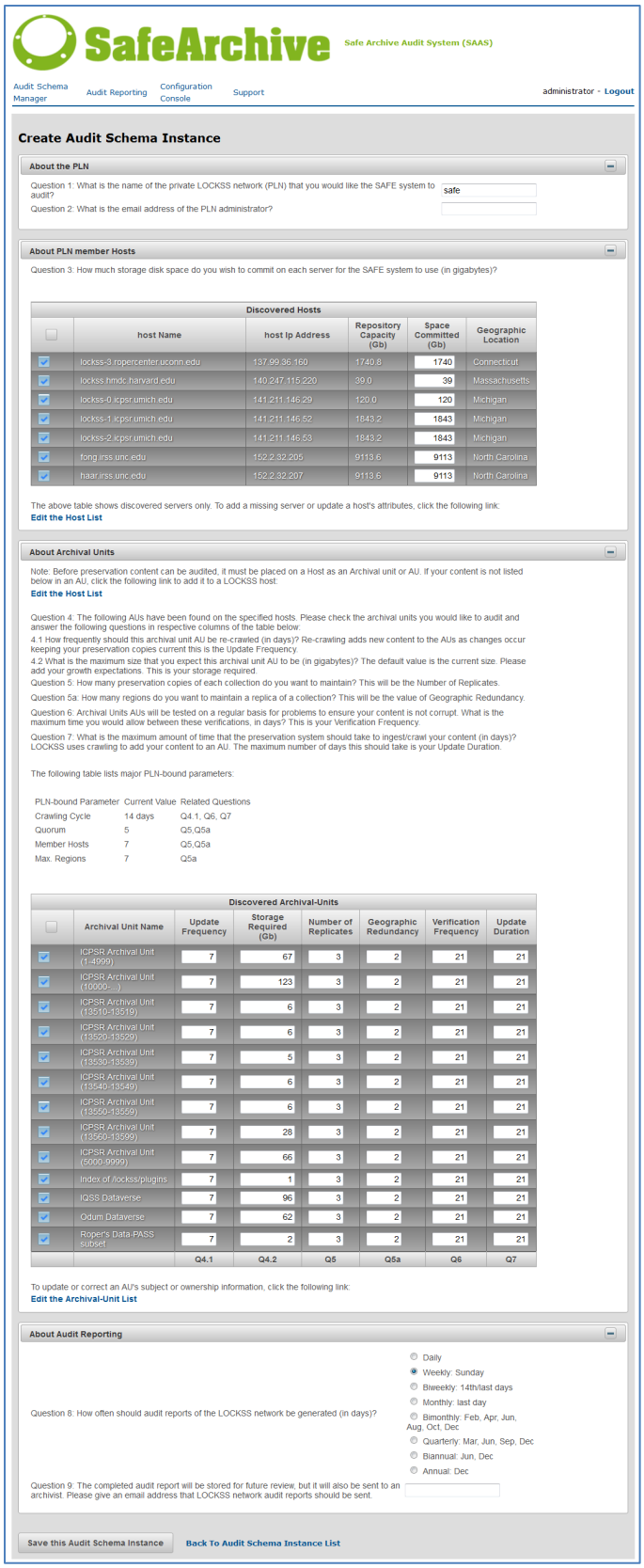
The process of creating an audit schema involves the consideration of several aspects of your PLN, hosts, and archival units. The SAAS presents several questions about each to assist with determining the audit policy appropriate for the collections in your network. The SAAS also offers a TRAC Checklist form, which stores your responses to TRAC criteria in a table. TRAC Audit reports indicate the degree to which the checklist has been completed. Follow the steps below to create network audit policy and to complete the TRAC Audit checklist.

---

1. From the Audit Schema Instance List page, click on the **Create a New Audit Schema Instance** button.



The Create Audit Schema Instance page contains a series of questions about your network. Each input field corresponds to a question that asks about a specific aspect of your audit policy. Enter your responses to each question in the appropriate fields. Click on the **Save the Audit Schema Instance** button to save your new audit schema.



SafeArchive — Safe Archive Audit System (SAAS)

Audit Schema Manager | Audit Reporting | Configuration Console | Support | administrator - Logout

Create 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? safe

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)?

Discovered Hosts

host Name	host Ip Address	Repository Capacity (Gb)	Space Committed (Gb)	Geographic Location
lockss-3.ropercenter.uconn.edu	137.99.36.160	1740.8	1740	Connecticut
lockss.hmdc.harvard.edu	140.247.115.220	39.0	39	Massachusetts
lockss-0.icpsr.umich.edu	141.211.146.29	120.0	120	Michigan
lockss-1.icpsr.umich.edu	141.211.146.52	1843.2	1843	Michigan
lockss-2.icpsr.umich.edu	141.211.146.53	1843.2	1843	Michigan
fong.irss.unc.edu	152.2.32.205	9113.6	9113	North Carolina
haar.irss.unc.edu	152.2.32.207	9113.6	9113	North Carolina

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.

The following table lists major PLN-bound parameters:

PLN-bound Parameter	Current Value	Related Questions
Crawling Cycle	14 days	Q4.1, Q6, Q7
Quorum	5	Q5,Q5a
Member Hosts	7	Q5,Q5a
Max. Regions	7	Q5a

Discovered Archival-Units

Archival Unit Name	Update Frequency	Storage Required (Gb)	Number of Replicates	Geographic Redundancy	Verification Frequency	Update Duration
ICPSR Archival Unit (1-4999)	7	67	3	2	21	21
ICPSR Archival Unit (10000-...)	7	123	3	2	21	21
ICPSR Archival Unit (13510-13519)	7	6	3	2	21	21
ICPSR Archival Unit (13520-13529)	7	6	3	2	21	21
ICPSR Archival Unit (13530-13539)	7	5	3	2	21	21
ICPSR Archival Unit (13540-13549)	7	6	3	2	21	21
ICPSR Archival Unit (13550-13559)	7	6	3	2	21	21
ICPSR Archival Unit (13560-13599)	7	28	3	2	21	21
ICPSR Archival Unit (5000-9999)	7	66	3	2	21	21
Index of /lockss/plugins	7	1	3	2	21	21
IQSS Dataverse	7	96	3	2	21	21
Odum Dataverse	7	62	3	2	21	21
Roper's Data-PASS subset	7	2	3	2	21	21
	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

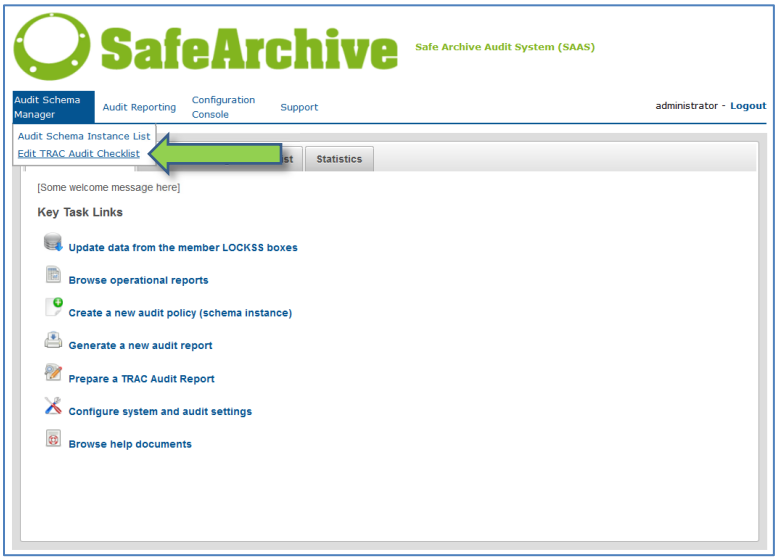
About Audit Reporting

Question 8: How often should audit reports of the LOCKSS network be generated (in days)?
- Daily
- Weekly: Sunday
- Biweekly: 14th/last days
- Monthly: last day
- Bimonthly: Feb, Apr, Jun, Aug, Oct, Dec
- Quarterly: Mar, Jun, Sep, Dec
- Biannual: Jun, Dec
- Annual: Dec

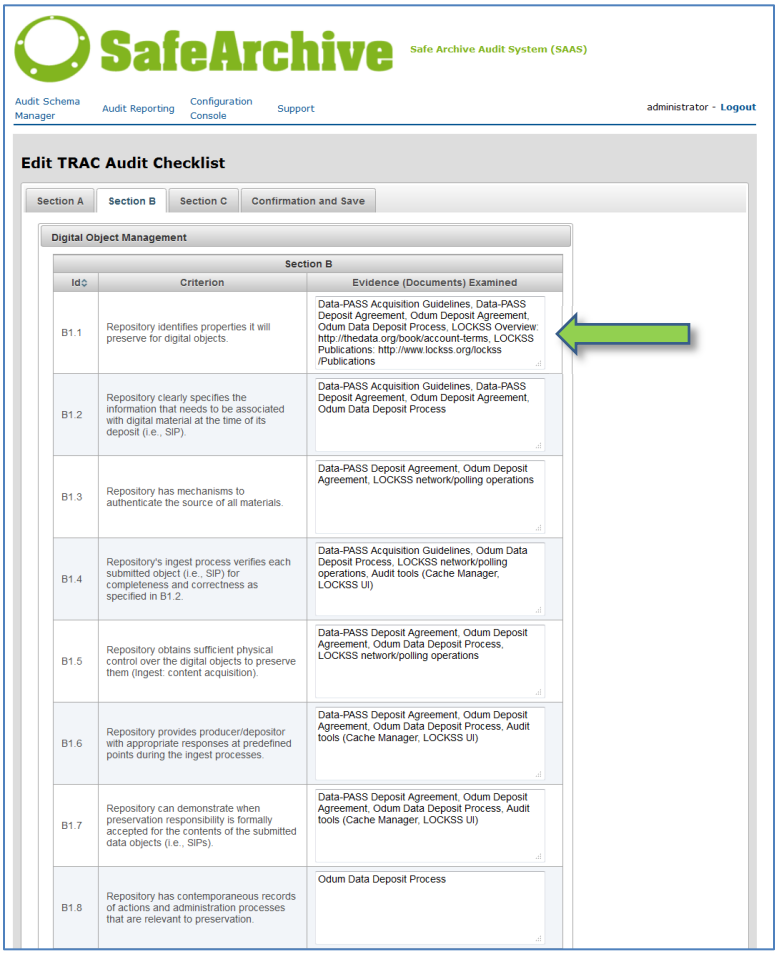
Question 9: The completed audit report will be stored for future review, but it will also be sent to an archivist. Please give an email address that LOCKSS network audit reports should be sent.

Save this Audit Schema Instance | Back To Audit Schema Instance List

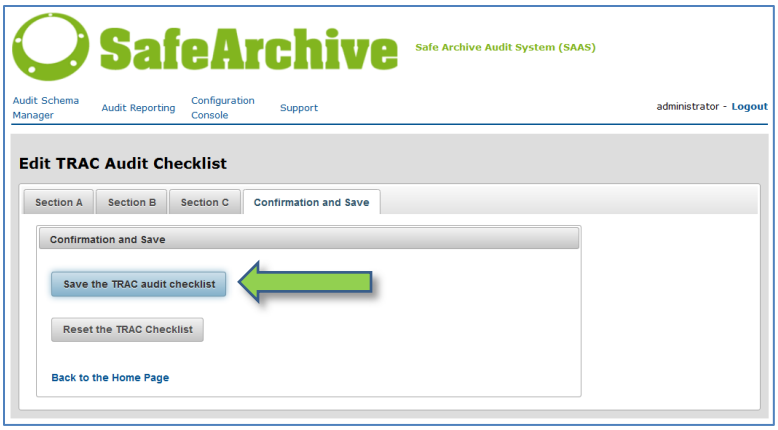
2. To generate a TRAC Audit report, you must first complete the TRAC Audit Checklist. To do this, click on the **Edit TRAC Audit Checklist** link in the Audit Schema Manager menu option



A table will display that includes tables for each of three sections of the checklist. Enter the evidence in the field that addresses each of the TRAC criterion. Click on the section tabs to fill in information for all three sections.



When you have entered data into each of the fields, click on the **Confirmation and Save** tab and click on the **Save the TRAC audit checklist** button.



Using the SafeArchive System

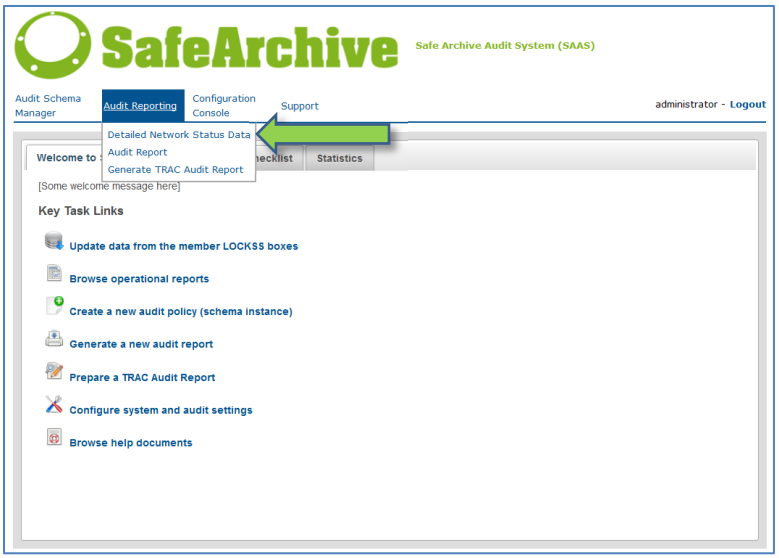
This section will enable you to complete the following tasks:

1. Viewing detailed network status data
2. Generating operational reports
3. Generating network audit reports
4. Generating TRAC audit reports

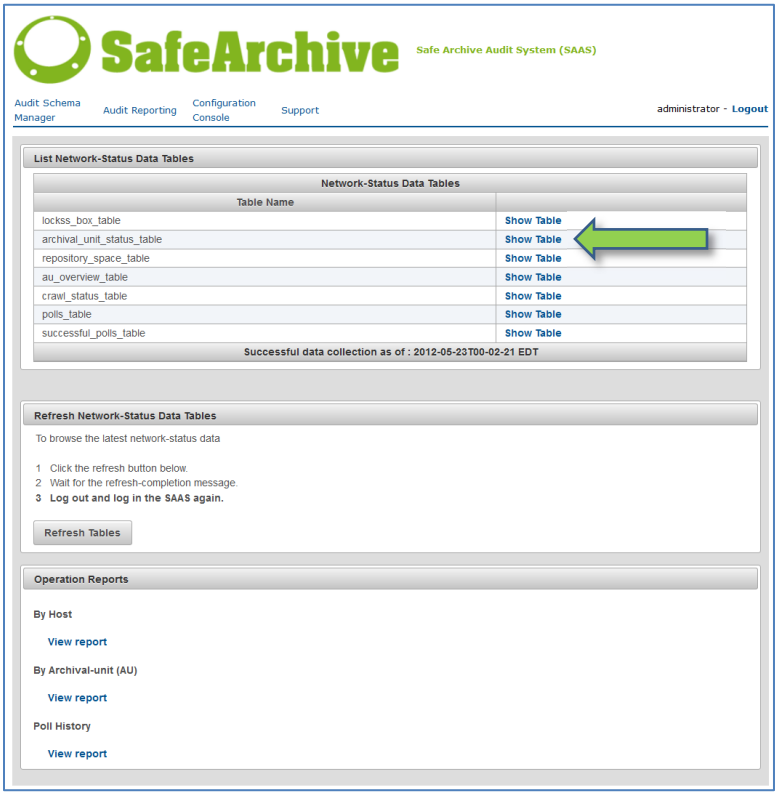
Once the SafeArchive System has been set up, you may view detailed network status data tables, generate audit and operational reports, and generate TRAC-audit reports. The instructions below provide guidance on accomplishing these reporting tasks.

---

1. From the main menu bar, navigate to the **Detailed Network Status Data** link in the **Audit Reporting** menu option.



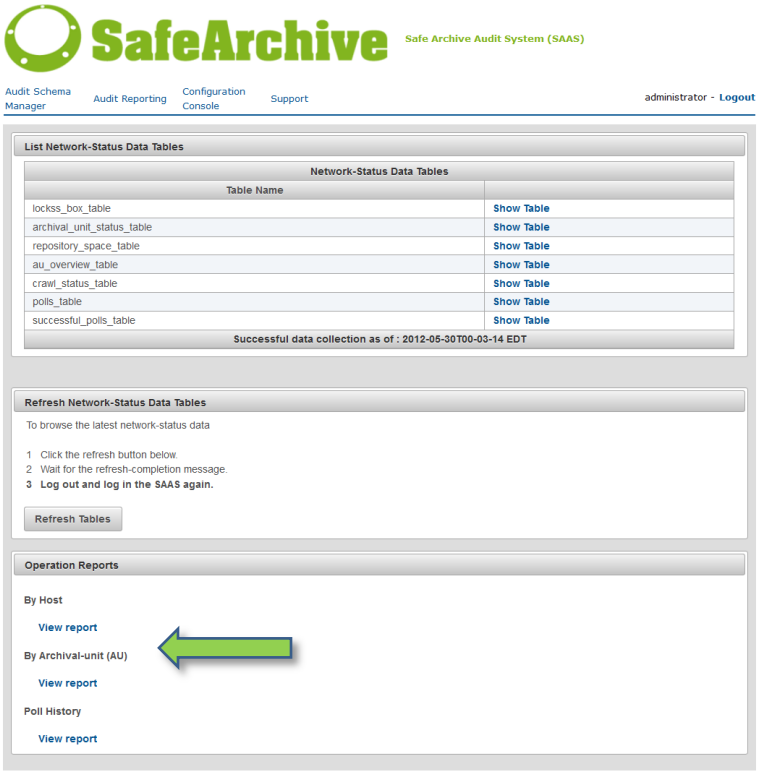
To view detailed status data tables, click on the **Show Table** link next to the table you wish to view.



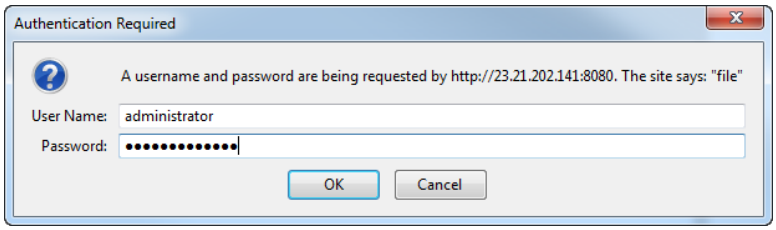
A table containing the network status data will appear. Clicking on the arrow buttons in the far left column will toggle a sub-table containing more detailed network table for the individual network host or archival unit.

---

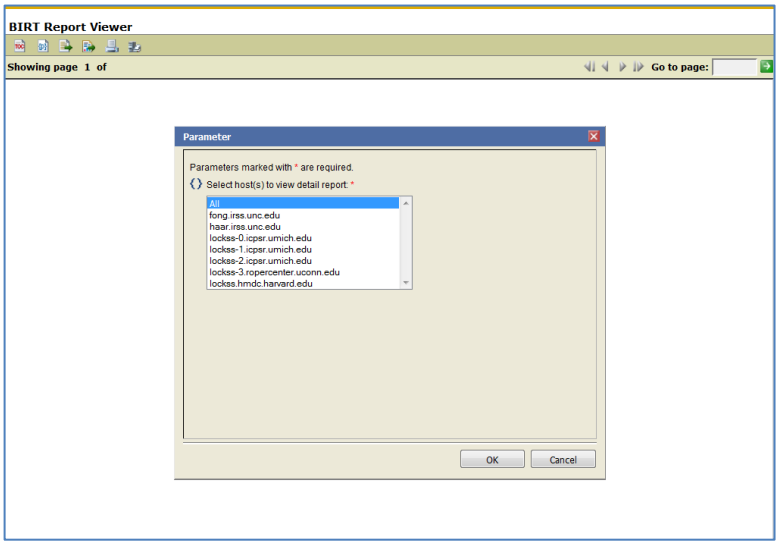
2. To generate printer-friendly operational reports via BIRT Report Viewer, return to the Detailed Network Status Data page. In the "Operation Reports" section at the lower part of the screen, click on the **View report** link either By Host or By Archival-unit (AU) as desired.



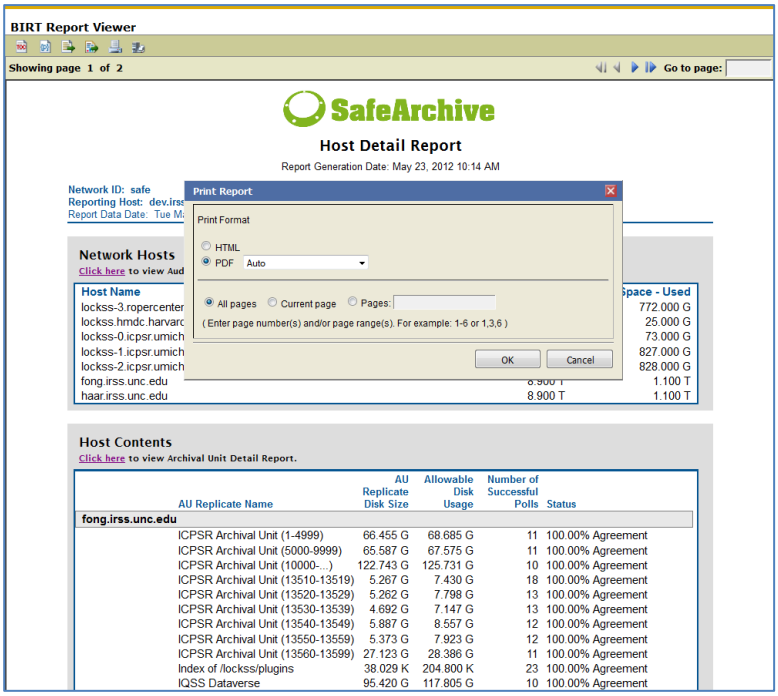
A dialogue box will appear that will prompt you to enter a User Name and Password. Enter the same user name and password used to access the SAAS. Click on the **OK** button.



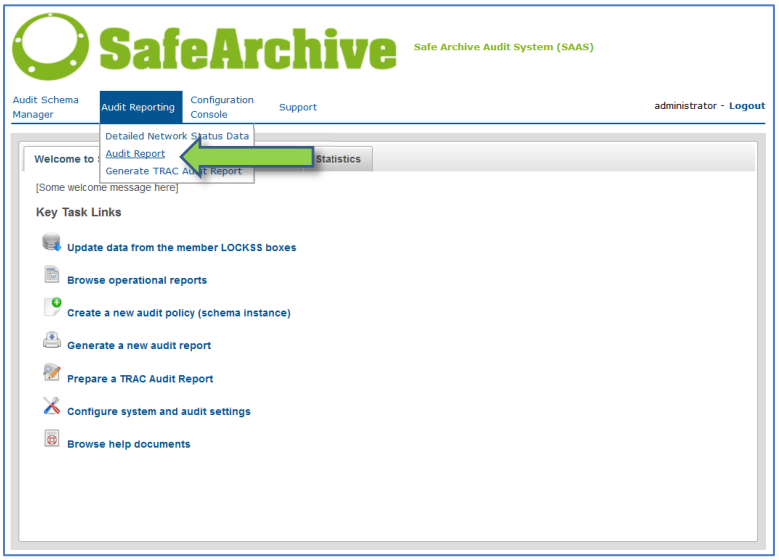
The BIRT Report Viewer will open a dialogue box requesting selection of the host/archival unit to be included in the report. Make your selection(s) and click on the **OK** button.



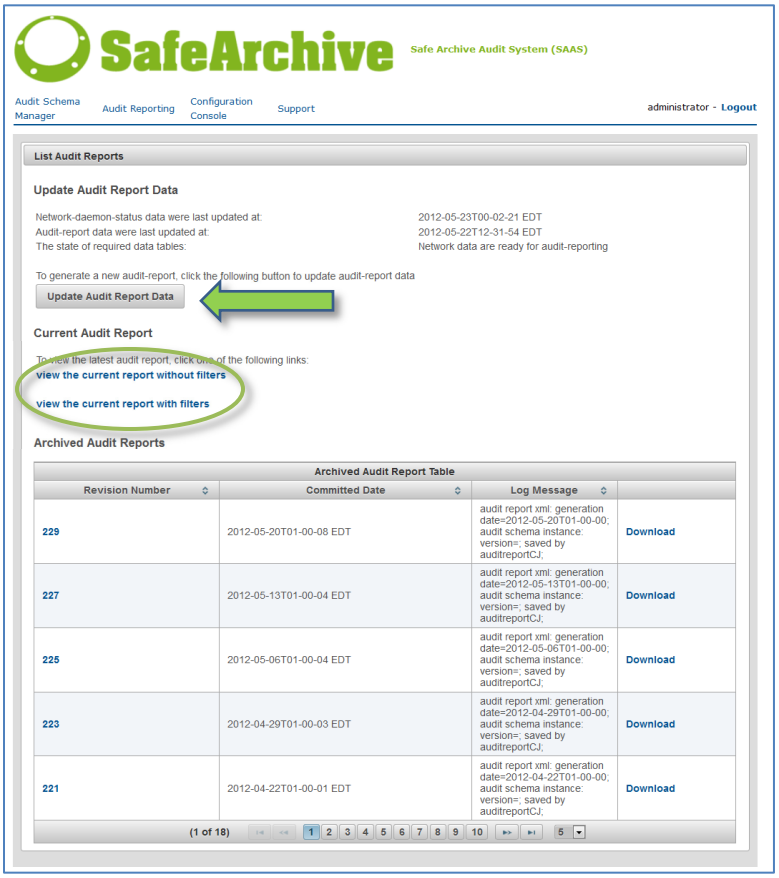
To print the report, click on the printer icon on the main menu bar at the top of the screen. A dialogue box will prompt you to select the print format and identify which pages of the report you wish to print. Click on the **OK** button to generate a printer-friendly version of the report.



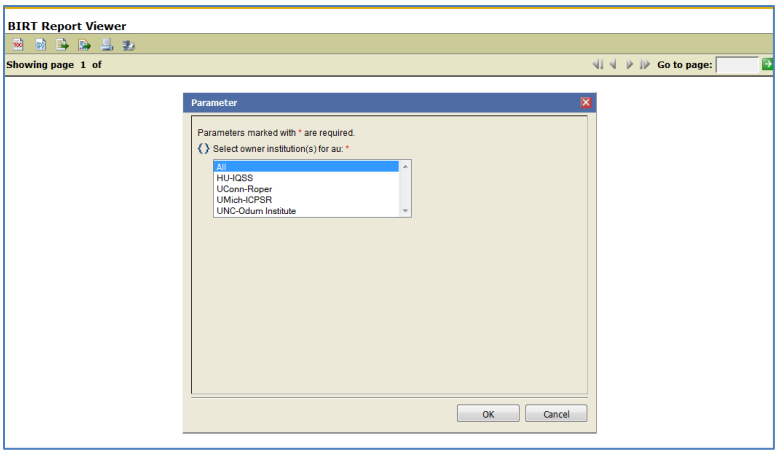
3. To generate an audit report, navigate to the **Audit Report** page in the Audit Reporting menu option at the main SAAS menu bar.



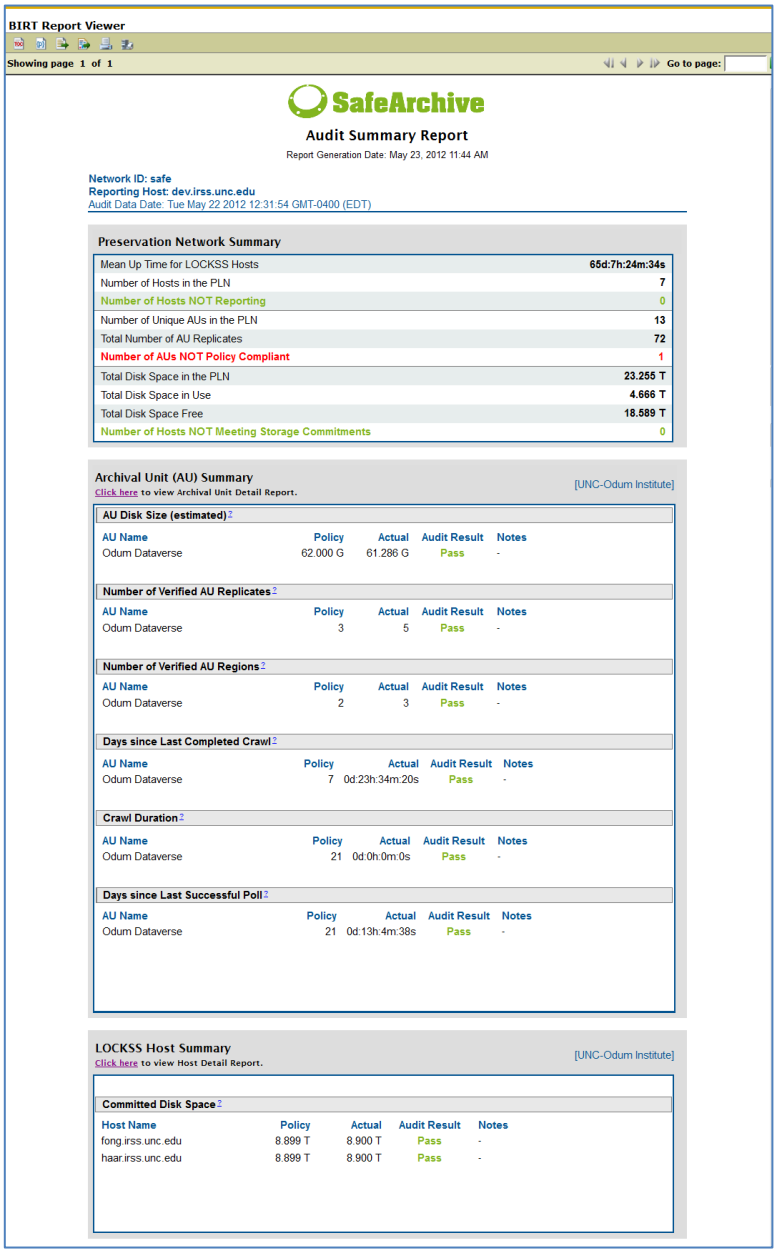
Click on the **Update Audit Report Data** button to fetch the most current network data. By clicking on the view the **current report without filters** link, the system will launch the BIRT report viewer displaying an audit report with details of the entire network.



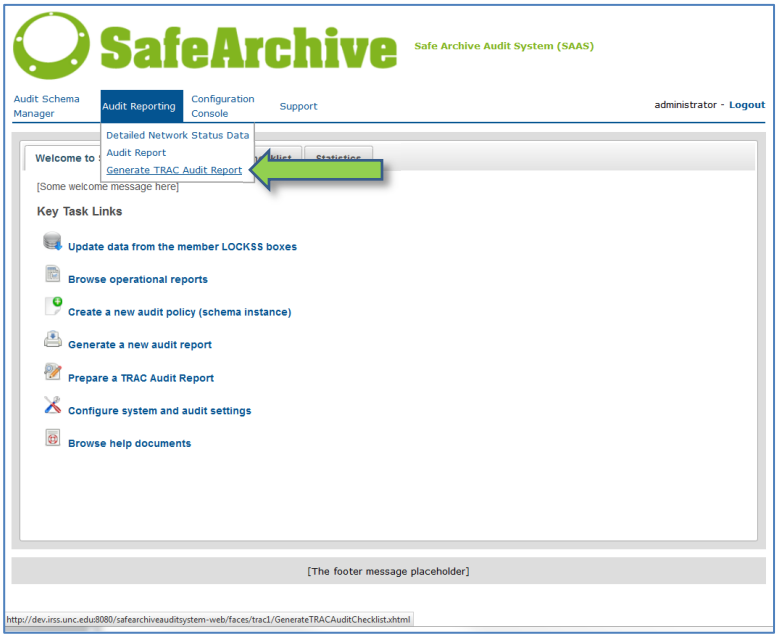
The **current report with filters** link displays the archival units and hosts owned by a specific institution based on your selection in the Parameter dialogue box.



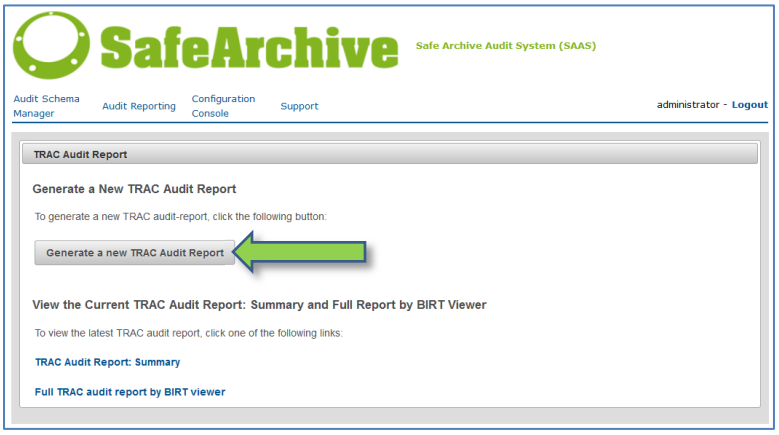
The BIRT Audit Report Summary provides audit results of the preservation network with pass/fail indicators. The Archival Unit Summary and Host Summary sections provide details of policy compliance or noncompliance for each individual archival unit and host.



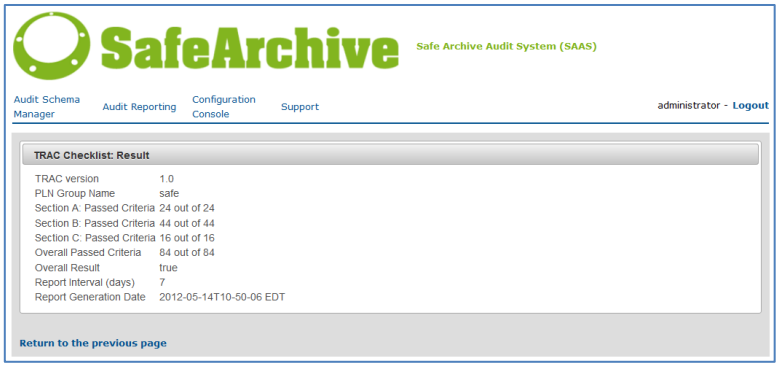
4. To view TRAC audit information, click on the **Generate TRAC Audit Report** link in the Audit Reporting menu option at the main SAAS menu.



After clicking on the **Generate a new TRAC Audit Report** button, click on either the **TRAC Audit Report Summary** or **Full TRAC audit report by BIRT viewer** link.



The TRAC Audit Report Summary link provides the completion status of the TRAC checklist. "Passed" status is given if text has been input into the TRAC checklist.



A printable BIRT report can also be generated by clicking on the **Full TRAC audit report by BIRT** viewer link on the **Generate TRAC Audit Report** page. This report not only displays the summary, but also it displays the TRAC checklist entries.

BIRT Report Viewer

Showing page 1 of 2 Go to page:

SafeArchive

Trusted Repositories Audit & Certification (TRAC) Report

May 23, 2012 12:29 PM

TRAC Audit Summary

PLN Group Name: safe
TRAC version: 1.0
TRAC Report data generation date: Mon May 14 10:50:06 EDT 2012

Complete
Section A: 24 out of 24 entered
Section B: 44 out of 44 entered
Section C: 16 out of 16 entered

TRAC Audit Checklist

Section A		Complete
A1.1	Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.	**Data-PASS About Page: http://www.icpsr.umich.edu/icpsrweb/DATAPASS/about.jsp#overview, Data-PASS Articles of Collaboration: http://www.icpsr.umich.edu/files/DATAPASS/pdf/collaboration.pdf**
A1.2	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.	**Data-PASS Articles of Collaboration**
A2.1	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.	**Data-PASS Articles of Collaboration, Odum Staff Page and Bios**
A2.2	Repository has the appropriate number of staff to support all functions and services.	**Data-PASS Publications and Presentations: http://www.icpsr.umich.edu/icpsrweb/DATAPASS /presentations.jsp, partner's Staff Pages**
A2.3	Repository has an active professional development program in place that provides staff with skills and expertise development opportunities.	**Data-PASS Publications and Presentations, Odum Internal Courses**
A3.1	Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.	**Data-PASS Articles of Collaboration, Data-PASS Confidentiality (pg. 6-7): http://www.icpsr.umich.edu/files /DATAPASS/pdf/confidentiality.pdf, Odum Mission Statement**
A3.2	Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.	**Data-PASS committee meetings, Odum Interwiki, Data-PASS Best Practices: http://www.icpsr.umich.edu /icpsrweb/DATAPASS/best-practices.jsp**

NOTE: To avoid additional charges, remember to TERMINATE YOUR INSTANCE! To do this, return to your AWS Management Console. Click on the **Instances** link in the main navigation menu at the left-hand side of the screen. Right-click on the SafeArchive instance. Click on **Terminate**.