

/Arkivum

Every bit archived

**EPrints Plugin V2.1:
Installation and User
Guide**

Part number	ARK/REPT/CUST/199
Revision	1.1A
Status	Draft
Creative Commons CC-BY	

Title: ePrints Plugin: Installation and User Guide

Part number: ARK/REPT/CUST/199

Revision 1.1A, Draft 12 Jan 2015



Copyright © 2015 Arkivum Limited.

This work is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

The content of this document is provided “as-is” and for informational use only. The information contained in this document is subject to change without notice, and should not be construed as a commitment by Arkivum. Arkivum assumes no responsibility or liability for any errors or inaccuracies that may appear in this document.

Except as permitted by such license, no part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, recording or otherwise, without the prior written permission of Arkivum Limited.

All other trademarks and trade names mentioned herein are hereby acknowledged and recognized as property of their respective owners.

Arkivum Limited
Langley Park Way
Langley Park
Chippenham
Wiltshire
SN15 1GE

Tel +44 1249 405060
info@arkivum.com

Document Approval

Reviewed by

Sign here using Acrobat
electronic signature

Approved by

Sign here using Acrobat
electronic signature

Change Control

Date	Version	Author	Description
4 th Dec 2013	1.0A	Mark Pembury	First draft
12 Jan 2015	1.1A	Matthew Addis	Update to change licensing to CC-BY

No change bars in first draft.

1. Contents

2. Background.....	2
3. The Arkivum A-Stor Service.....	3
3.1 Introduction.....	3
3.2 The Arkivum Plugin	5
4. Installation	6
4.1 Mounting the A-Stor Appliance	6
4.2 Create A-Stor data pools	8
4.3 Configure Apache	10
4.4 Perl dependencies	11
4.5 Installing the plugin	11
4.6 Configuring the plugin	12
4.6.1 <i>The mount directory.</i>	13
4.6.2 <i>The A-Stor appliance</i>	13
4.6.3 <i>The A-Stor scheduled job</i>	13
4.6.4 <i>The size threshold</i>	14
4.6.5 <i>The time estimate calculations</i>	15
4.6.6 <i>Configuring the plugin mode</i>	16
4.6.7 <i>Enable the plugin</i>	17
4.7 Check in E-Prints.....	17
5. User Guide	18
5.1 Archive Requests	18
5.2 Additional EPrint Actions	19
5.3 Request Archive Removal	20
5.4 Request Archive Restore	22
5.5 A-Stor Checker Task.....	24
5.6 Manage Records Screen	25
5.7 Manage Document Requests.....	26
5.8 Manage EPrint Requests.....	27
5.9 Authorizing a Request.....	28

2. Background

The Arkivum EPrints storage plugin allows content in an EPrints repository to be copied or moved into Arkivum's data archiving service.

Arkivum provides online data archiving as a service to organizations that need to keep data for long-term compliance or reuse.

The company is a spin out of the University of Southampton and is based on a decade of expertise and technology created from working with large archives on digital preservation and long-term data retention and access. As part of the Arkivum service, three copies of customer data are held in three UK locations: two online in secure data centers and one offline in escrow with a third-party. Arkivum actively manages data integrity through regular checks and migrates media and infrastructure to counter obsolescence and to ensure prices remain low.

Good preservation practice, trained staff and very carefully controlled processes means that Arkivum can offer a guarantee of data integrity.

All data returned from our service is always bit-for-bit identical to the data the customer supplied, with no restrictions on time or volume. The guarantee is backed by worldwide insurance and is included in our contract and SLA. Arkivum is certified to ISO27001 and is regularly audited for the integrity, confidentiality and availability of data assets in our possession. This approach makes the Arkivum service ideal for long-term storage of research data, for example to meet research council retention requirements and to ensure it is available for easy access in the future.

For more information see www.arkivum.com including our series of articles and webinars on Research Data Management.

3. The Arkivum A-Stor Service

3.1 Introduction

This document identifies some of the interactions needed between EPrints and the Arkivum service. To understand these, we start by looking at what happens when files are ingested into the A-Stor service and then subsequently retrieved.

When files are added to the A-Stor service, e.g. by a copy into the file share provided by the A-Stor appliance, several steps will happen internally within the appliance and within the A-Stor service.

1. The file is stored on the local appliance hard disk
2. Checksums are generated for the received file (MD5, Alder32).
3. The file is encrypted and has the encryption verified.
4. The encrypted file is replicated to the two Arkivum data centres.
5. The encrypted file is written along with other customer files to a data tape that is stored in escrow.

When the file first gets copied to the appliance it is labelled as in a 'RED' state. This means there is a local copy of the file on the appliance, but the file has yet to complete ingest and replication processes. The file is also 'locked' at this point, i.e. deletion or updates to the file are prevented – the file becomes read only.

When the file has been replicated to the data centre and successfully gone into escrow then the file is labelled as in a 'GREEN' state. Only when the file is green will it be considered as safely stored in the service and deletion of the appliance copy is enabled.

The local copy of the file on the appliance can be removed if required, for example to free up space, but it can also be kept on the appliance 'in cache' if there is likely to be the need to access the file in the near future.

The Arkivum data integrity guarantee does not start until:

- (a) The file is confirmed to be in the green state, which includes confirmation that it has been received into escrow (not just dispatched to escrow).
- (b) The customer has agreed that the file has been correctly received by A-Stor, i.e. they accept that the checksums generated on the file are correct.

The first part is needed from a data safety perspective, i.e. to ensure no guarantee is made until there are multiple copies of the file in the service. The second part is

needed from an error trapping and non-repudiation perspective, i.e. it allows a chain of custody to be established between the customer and Arkivum that shows that the data has been correctly received.

Files in A-Stor are WORM by default (write once, read many) and attempts to delete the file over the file share are blocked. If a file does need to be deleted then a request to do so can be made over the A-Stor web interface. Deleting a file means that first A-Stor will no longer make the file accessible to the user through the appliance. This is instantaneous. Physical deletion of the data in the service happens next and can take some time, for example because the escrow copy has to be deleted.

The upshot of the way that A-Stor works as described above means that EPrints will need to interact with A-Stor in a way that involves more than simple file system operations.

When files are copied into A-Stor:

- (a) Checksums as generated by A-Stor should be checked against checksums generated from files on the source file system (e.g. local EPrints storage).
- (b) The copy in A-Stor should not be considered 'safe' until after the files have gone into the green state.

Going from red to green can take several days (or longer) due to the need to deposit an offline copy of the data with the escrow provider. Therefore, an asynchronous interaction process is needed between A-Stor and EPrints. This could be implemented in various ways, e.g. call-backs from A-Stor to EPrints when there is an update on the status of a file, or by EPrints polling A-Stor to check the status of a file on a period basis.

When files need to be retrieved from A-Stor:

- (a) A-Stor should be checked to see if the file is still in the appliance cache or not.
- (b) If there is not a copy in the appliance cache, then a restore to cache may need to be triggered in advance of the data being served through EPrints.

If the file size is large then retrieval from the Arkivum service can take some time and has a 'cost' in terms of using network bandwidth that the University might otherwise want to use for something else. Therefore, restore to cache requests to A-Stor may need to be scheduled so the data transfer takes place at times of low network activity, e.g. overnight.

When files need to be deleted from A-Stor:

- (a) A deletion request has to be made through a secure channel (e.g. admin API or web interface)

-
- (b) The deletion should not be considered as ‘complete’ until confirmation by A-Stor.

3.2 The Arkivum Plugin

This text will take you through the process of installing and configuring the Arkivum Plugin which will then allow EPrints to store its content in the Arkivum A-Stor service. The plugin will use the A-Stor appliance as a storage cache mounted locally via Samba (CIFS/Windows) share to store that data.

The plugin uses EPrints event tasks to manage the process of copying, restoring and removing data from the A-Stor service as a background task.

The process of copying or removing data from the A-Stor service can happen in one of two ways.

1. **Manual request to archive a document attached to an EPrint item.** Individual documents within an EPrint can be copied to the archive on request through the EPrints user interface. This requires an EPrints administrative user to choose which documents to copy. The local copy in EPrints is retained and not deleted, i.e. this mode will cause an additional copy of the document to be held in A-Stor. Or by a request / approval process which allows the user to archive a document for a specific EPrint item.
2. **Automatic archiving of EPrints when they are submitted to the repository.** The plugin can use an EPrints database trigger to automate the movement of an entire EPrint to the A-Stor archive. The EPrint is first copied to A-Stor when the EPrint is ‘moved to the repository’ through the EPrints user interface. The local copy of the EPrint is then deleted when the EPrint has been confirmed as safely stored in A-Stor. This process happens automatically.

The first version of the plugin introduced the first option.

The second version of the plugin builds on that functionality by allowing EPrints to automate the submission of an EPrint to the A-Stor service when it is moved into a repository.

The plugin can be configured to operate in mode 1 only or to operate in mode2+mode1.

4. Installation

The A-Stor service is accessed from EPrints using the SMB protocol (also known as CIFS) for ingest and access to files and through a REST API over https for monitoring and management. A-Stor exposes the REST API and file system interfaces through an appliance that is installed at the customer site. This uses Samba to expose a network file share. This means that the A-Stor network share needs to be mounted on the server running EPrints so that the plugin to EPrints can copy files to the A-Stor network share and read files from the A-Stor network share.

The installation process for the network share is:

- Mount the SMB network share on the EPrints server
- Check permissions are set properly so EPrints can read and write files to that share
- Install and configure the plugin
- Test that the plugin is visible and functioning correctly within EPrints.

4.1 Mounting the A-Stor Appliance

Install smbfs on the EPrints server (if it is not already installed)

For Debian this would be:

```
apt-get install smbfs
```

For Redhat / CentOS this would be:

```
yum install cifs-utils
```

Note: Red Hat / CentOS uses CIFS mounting rather than samba (so replace SMBFS for CIFS if using Red Hat)

To configure the mount point:

- Make sure you have read / write access to /etc/fstab
- Determine the UID of EPrints (likely to be EPrints) and the GID of the webserver group (likely to be www-data, EPrints or apache)
- Create a directory and read/write access for the above user somewhere on the system that will be the mount point for the share.

Note: The file system exposed by the A-Stor appliance does not support SELinux. On Redhat or CentOS this means either configuring SELinux appropriately or switching SELinux off.

To set-up the mount point for the A-Stor appliance you will need to do the following:

- Create the directory where you want to mount the Astor Archive on to. For example, if we wanted to mount the appliance to a directory called `/mnt/arkivum`:

```
$ mkdir /mnt/arkivum
```

- Edit `/etc/fstab` and add line to configure the mount point. For example:

```
//ASTOR_SERVER/astor /mnt/arkivum cifs
defaults,guest,file_mode=0666,dir_mode=0777,uid=USERS_UID,gid=USERS_GID,forcegid,forceuid,rw 0 1
```

If you want to connect as a specific domain user, use

```
//ASTOR_SERVER/astor /mnt/arkivum cifs
defaults,sec=none,rw,soft,username=DOMAIN_USERNAME,password=DOMAIN_PASSWORD,dom=DOMAIN,uid=EPRINTS_UID,gid=APACHE_GID,forcegid,forceuid
0 0
```

- Mount the A-Stor file system:

```
$ mount /mnt/arkivum
```

- Check to see if the mount has succeeded by running the command:

```
$ df -h
Filesystem              Size  Used Avail Use% Mounted on
/dev/mapper/VolGroup00-LogVol100
                        16G  4.0G   11G  27% /
/dev/hda1                99M   13M   81M  14% /boot
tmpfs                   1004M    0 1004M   0% /dev/shm
//172.18.6.9/astor      7.5T  1.8T  5.7T  24% /mnt/arkivum
```

Note: If you are running Red Hat or CentOS then you need to check if you are running Security Enhanced or SELinux. The mounted A-Stor file system does not support SELinux so you either need to disable it or ensure that the EPrints / Apache user has the appropriate read / write permissions.

Once the A-Stor file system is successfully mounted on the EPrints server any folders or files written to the mount will be stored in the A-Stor archive service.

Note: In the above example, if the A-Stor appliance server was rebooted then the mount point would become unavailable. You would then need to unmount and mount the A-Stor appliance before EPrints could write to the mounted file system again.

4.2 Create A-Stor data pools

You should create a data pool in A-Stor for the EPrints repository that you will be using for the plugin. This needs to be done before installing and using the plugin.

Using the A-Stor Web UI for the appliance do the following:

1. Create a directory in the root of the A-Stor file system with a name that matches the repository that you want to archive. The directory name should be the same as the one created by EPrints for local storage, e.g. as found in `/usr/share/EPrints3/archive` or `/opt/EPrints3/archive`.
2. Create a data pool for that directory. It is simplest if the data pool name matches the repository name in EPrints. The directory name for the data pool must match the directory created in Step 1 above. Choose for the data pool to be unencrypted and have an escrow threshold of 1.2TB.
3. Test access to the data pool from the EPrints server.
 - a. As the EPrints user, check from the command line that the data pool is visible in the mounted A-Stor directory.
 - b. As the EPrints user, check that the permissions are set correctly to allow creation of a subdirectories in the data pool, e.g. navigate to the new data pool and execute `'mkdir -p testdir1/testdir2/testdir3'` and check the folder hierarchy has been created (it can then be deleted).

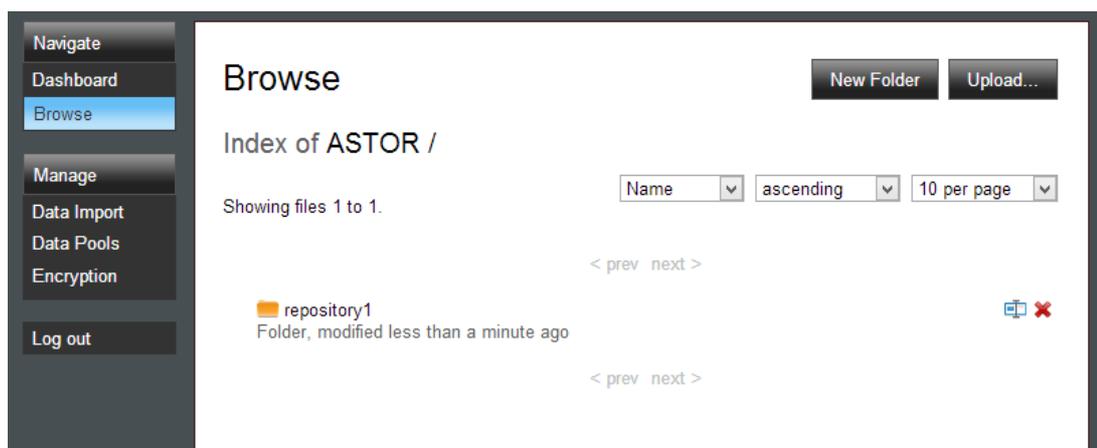


Figure 1 Create a directory (using the New Folder button) for the repository to be archived. The directory in this examples is called “repository1”

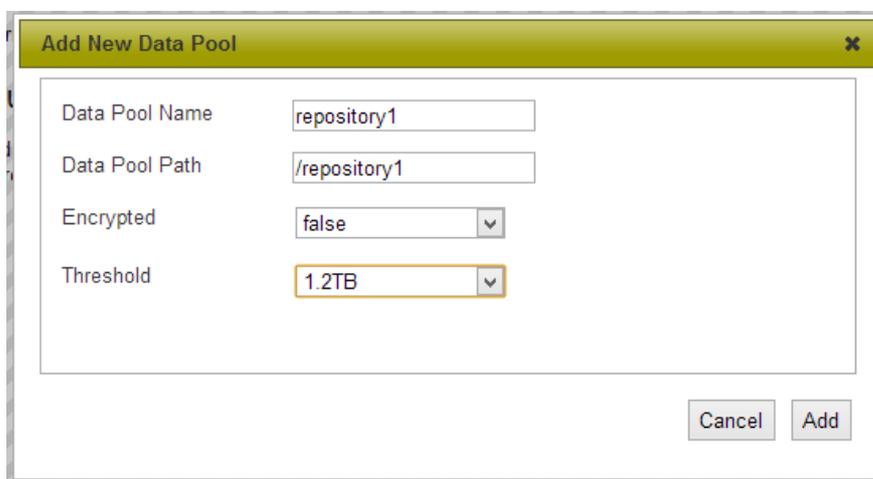


Figure 2 Create a data pool for the repository (using the Data Pools button on the left hand side of the A-Stor menu). The Data Pool path should match the directory created in the A-Stor filesystem. The path should include a leading /.

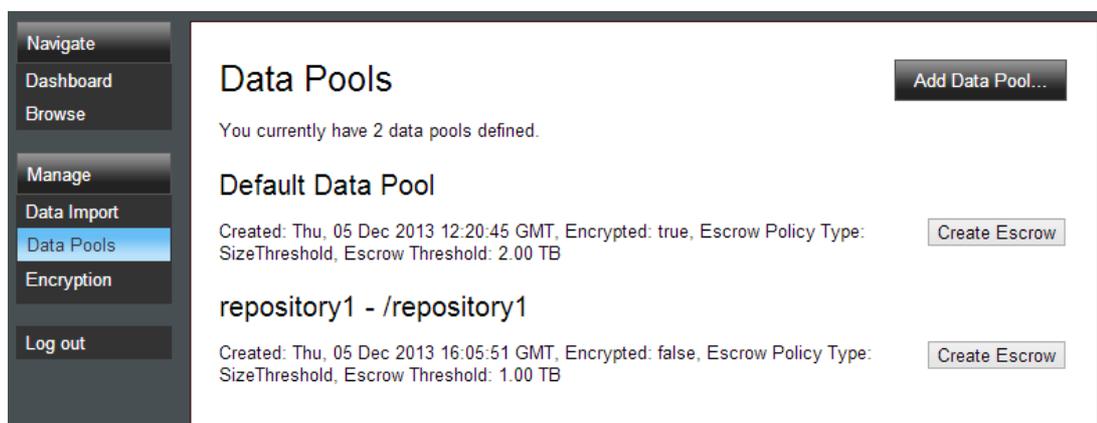


Figure 3 The data pool for EPrints (repository1 in this example) can be seen in the list of A-Stor data pools.

4.3 Configure Apache

Now that we have the A-Stor appliance mounted on the EPrints server we need to ensure that the Apache Web Server can access the files stored in that folder.

To do this you will need to add the following configuration for your mount folder to the apache configuration file for EPrints.

```
<Directory "YOURDIR_ON_EPRINTS">
  Options FollowSymLinks
  AllowOverride none
  Order allow,deny
  Allow from all
</Directory>
<Location "YOURDIR_ON_EPRINTS">
  PerlSetVar EPrints_ArchiveID fluffy
  Options +ExecCGI
  Order allow,deny
  Allow from all
</Location>
```

The apache configuration file will be located in one of the following directories depending on whether you are running http or https.

```
%eprints%/cfg/apache
%eprints%/cfg/apache_ssl
```

Within the appropriate folder you will find an include file for your EPrints repository. An example configuration file for one of the test EPrints servers is show below:

```
<VirtualHost *:80>
  ServerName dcleprints.test.local

  ServerAdmin mark.pembury@arkivum.com

  <Directory "/mnt/arkivum/dcleprints/">
    Options FollowSymLinks
    AllowOverride none
    Order allow,deny
    Allow from all
  </Directory>

  <Location "">
    PerlSetVar EPrints_ArchiveID dcleprints
```

```
Options +ExecCGI
Order allow,deny
Allow from all
</Location>

<Location "/mnt/testr320/dcleprints/">
    PerlSetVar EPrints_ArchiveID dcleprints
    Options +ExecCGI
    Order allow,deny
    Allow from all
</Location>

# Note that PerlTransHandler can't go inside
# a "Location" block as it occurs before the
# Location is known.
PerlTransHandler +EPrints::Apache::Rewrite
</VirtualHost>
```

4.4 Perl dependencies

The plugin has the following Perl module dependencies:

- Data::Dumper
- DateTime::Format::ISO8601
- File::Basename
- IO::Socket::SSL
- JSON
- LWP::UserAgent

Typically the following modules may need to be explicitly installed:

- DateTime::Format::ISO8601
- IO::Socket::SSL
- JSON

There is an additional Perl script available that will check if Perl can communicate as required with the A-Stor Appliance via the REST interface used in the Plugin.

4.5 Installing the plugin

The Arkivum A-Stor EPrint plugin is distributed as an epm file. To install this file you will need to do the following with admin rights in EPrints:

-
- Use a web browser to navigate to you EPrints server
 - Login to the admin section of EPrints
 - Click on the System Tools tab
 - Click on the EPrints Bazaar button
 - Click on the Available tab and navigate to the bottom of the screen
 - Click on the choose file button and locate the Arkivum-2.0.epm installation file.
 - Click on the install button

The plugin will install and then attempt to enable the plugin. The plugin will check to see that the required Perl dependencies are installed on the system. If they are then the plugin will be enabled.

Before we use the plugin we will need to disable it and set-up the configuration options. By default, the main configuration options are set to be blank.

4.6 Configuring the plugin

Once the plugin is installed and before it can be used, there are three configuration settings that need to be made. You will need to change these configuration settings for the repository that you've installed the plugin for.

These tell the plugin where the mounted A-Stor appliance is located and how it can communicate using its REST API.

The default configuration file for the plugin can be found in:

```
%eprints%/lib/epm/Arkivum/cfg/cfg.d/x_arkivum.pl
```

You will need to edit this file if you want to change the default settings that are used when you enable the plugin.

The repository configuration file for the plugin can be found in:

```
%eprints%/archive/%repo%/cfg/cfg.d/x_arkivum.pl
```

You will need to edit this configuration file to change the required settings if you do not want to edit the default configuration file and re-enabled. If you edit this configuration then you will need to reload your EPrints configuration.

4.6.1 The mount directory.

This setting indicate where the mounted directory of the A-Stor appliance is located.

```
$c->{plugins}->{"Storage::ArkivumStorage"}->{params}->{mount_path} =  
"";
```

If we had previously set-up the mount directory as /mnt/arkivum then this configuration settings would become:

```
$c->{plugins}->{"Storage::ArkivumStorage"}->{params}->{mount_path} =  
"/mnt/arkivum";
```

4.6.2 The A-Stor appliance

These two settings tell the respective plugins the URL to use to communicate with the A-Stor appliance:

```
$c->{plugins}->{"Storage::ArkivumStorage"}->{params}->{server_url} =  
"";  
$c->{plugins}->{"Event::Arkivum"}->{params}->{server_url} =  
"";
```

If we had set-up the appliance to use an IP address of 172.18.6.9 then the configuration settings would become:

```
$c->{plugins}->{"Storage::ArkivumStorage"}->{params}->{server_url} =  
"https://172.18.6.9:8443";  
$c->{plugins}->{"Event::Arkivum"}->{params}->{server_url} =  
"https://172.18.6.9:8443";
```

Note: You can also use a hostname but you must specify the port that is used to connect to the A-Stor appliance. You can test the URL for the appliance in web browser.

Once the settings in section 3.5.1 and 3.5.2 are set-up you can now enable and test the plugin. You may however want to change some of the other settings.

4.6.3 The A-Stor scheduled job

When the plugin is installed it will add a scheduled task that is executed by the indexer. This task will check EPrints for a request by EPrints to perform an action with the A-Stor service.

By default the schedule is set-up to run the checker task every 15 minutes. This can be changed by editing the following file:

```
/opt/eprints3/lib/plugins/EPrints/Plugin/Screen/EPMC/Arkivum.pm
```

You will then need to change the following section

```
EPrints::DataObj::EventQueue->create_unique( $repo, {  
    pluginid => "Event",  
    action => "cron",  
    params => ["0,15,30,45,60 * * * *",  
              "Event::Arkivum",  
              "astor_checker",  
    ],  
});
```

The default specifies that the checker task is to run one the hour, 15 minutes past, etc. Please refer to the cron manual page on your Linux system for further information regarding how you specify schedules.

4.6.4 The size threshold

Once an EPrint item has been archived in the A-Stor service then its local copy in EPrints is removed (mode 2 of the plugin). The EPrint is still available to users at any time through the EPrints server, which access the files via the local file system mount from the A-Stor appliance.

The A-Stor appliance provides a cache so that files can be held on the appliance as well as stored in the data centers and on Escrow tape. If the appliance requires space then it will automatically remove any archived files with the knowledge that it can be read directly from the data center and restored to the appliance if required.

Files can also be specifically requested to be held in the cache or restored to the cache via the REST API.

Therefore, over time, some files will have copies in both the A-Stor cache and the Arkivum data centers and some files will only have copies in the data centers.

When a user wants to access a file through EPrints then the plugin will test to see if it is in the appliance cache. If it is then the file will be served to the user as normal.

If the file is not in the cache and is small (below a threshold size) then it will be automatically retrieved from the data center and restored to the cache for access.

If the file is not in the cache and is large (above the threshold size) then the request from the user to access the file will go through a review/approve process.

The threshold that determines if a file can be served directly from the A-Stor service or if the user needs to be redirected to a request screen so that the restore can be approved is set in the plugin configuration. The default value is set as 1GB in bytes.

```
# The size threshold (in bytes, default is 1GB = 1024 x 1024 x 1024)
# used to determine if we
# server a file directly from the data center if it is not stored on
# the local appliance
# If the file is less than or equal to this threshold it will be
# served from the data center
# directly otherwise the user is redirected to an eprint screen to
# request a restore.

$c->{plugins}->{"Storage::ArkivumStorage"}->{params}-
>{redirect_threshold} = 1073741824;
```

If a user requests a file from an EPrints that has been archived and the following conditions are met then they will be redirected to a restore request screen:

- File has been archived in A-Store (green)
- File does not have a copy stored on the A-Stor appliance
- File size is greater than 1GB

If the conditions are met then the user will be redirected to a request screen where they can request a restore of the files that belong to the EPrint. If the request is approved they will be notified via email when the file is available.

If these conditions are not met then the file will be delivered from the A-Stor service and automatically restored to the appliance if required.

4.6.5 The time estimate calculations

When a request to restore an EPrint is made the request screen will attempt to calculate how long this is likely to take. The following configuration options are used to make that time estimate.

```
# These two parameters are used to estimate the time to restore an #
# eprint to the local A-Stor cache.
# Factor_a is multiplied by the total size of
# the ePrint (in bytes) which is then added to factor_b.
#
# Both parameters are expected to be numbers in seconds
```

```
# Time to restore 1MB (1024 x 1024) in seconds
$c->{plugins}->{"Screen::EPrint::AStorRestore"}->{params}-
>{factor_a} = 2;

# Specify an expect wait time in seconds
$c->{plugins}->{"Screen::EPrint::AStorRestore"}->{params}-
>{factor_b} = 900;
```

4.6.6 Configuring the plugin mode

As highlighted in Section 3.2 this plugin builds on work done previously in version 1.0. In version one a user could manually submit a request for a document to be archived in A-Stor and if the request was approved then it would be archived.

In version 2.0 of the plugin this has been extended to apply to the EPrint rather than just a document and the request to archive an EPrint is made automatically.

The plugin for version 2.0 is preconfigured to run in the automatic mode. If you wish to run the version 1.0 manual mode then you will need to change the plugin configuration file. This will prevent the plugin from automatically submitting a request to archive the EPrint when it is moved into the repository.

To run the plugin as per version 1.0 with no automatic submission, you will need to make the following changes to the plugin configuration file:

1. Enable the document request screen by changing the following line in the configuration file. By default the value is set to 1 to disable the screen.

```
$c->{plugins>{"Screen::EPrint::Document::AStor"}{params}{disable} = 0;
```

2. Disable the database trigger that is responsible for the auto submission by either commenting out or removing the following lines in the configuration file:

```
#$c->add_dataset_trigger( "eprint", EP_TRIGGER_STATUS_CHANGE ,
#   sub
#   {
#       my ( %params ) = @_;
#       my $repository = %params->{repository};
#
#       return undef if (!defined $repository)
#       . . .
#   }
```

```
#) ;
```

4.6.7 Enable the plugin

Once you have made the required changes to the plugin you may need to re-save the plugin and generate a new checksum before you can enable it. To do this use the following steps:

- Use a web browser to navigate to your EPrints server
- Login to the admin section of EPrints
- Click on the System Tools tab
- Click on the EPrints Bazaar button
- Click on the Developers tools tab
- Click on the Edit button of the Arkivum plugin
- Click on the Save and Return button
- Click on the Installed tab
- Click on the Enabled button.

If all goes well you should see a message indicating that the plugin has been enabled.

Note that it is important to use the developer tools to save the plugin before it is enabled. It is not sufficient just to enable the plugin or reload the EPrints configuration.

4.7 Check in E-Prints

If the installation is successful and the plugin has been enabled correctly, you should see the main event astor-checker task here:

```
E-prints web page --> LOGIN --> ADMIN --> SYSTEM TOOLS --> STATUS --> Background Task Queue
```

This astor-checker should appear similar to this:

Status	Start Time	Plugin	Action	Parameters	
Waiting	5 December 2013 09:32:00 UTC	Event	cron	['****', 'Event::Arkivum', 'astor_checker']	

There are also several additional screens that will become available when required. Please refer to the user guide section for further information.

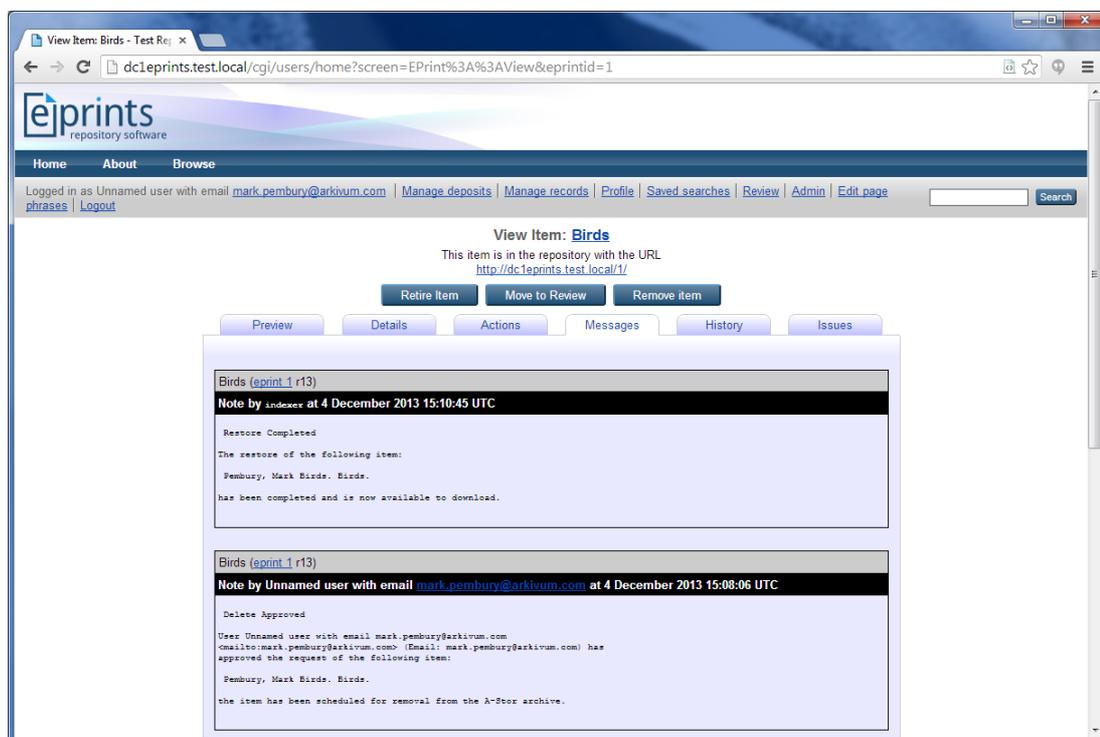
5. User Guide

5.1 Archive Requests

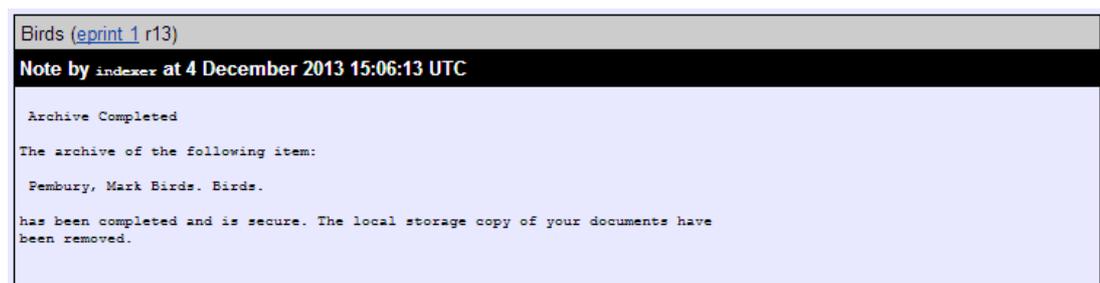
In version 2.0 of the plugin a request to archive an EPrint item in A-Stor is made everytime a new item is added to the repository.

This request will then get processed by the indexer and all of the documents (and files) that are attached the EPrint are then copied on the A-Stor appliance. From the user perspective no other interaction is required.

Once the EPrint has been archived a note is added to the history of the EPrint.



The messages tab of an EPrint showing the notes from the A-Stor plugin



The note added to the EPrint once it has been archived.

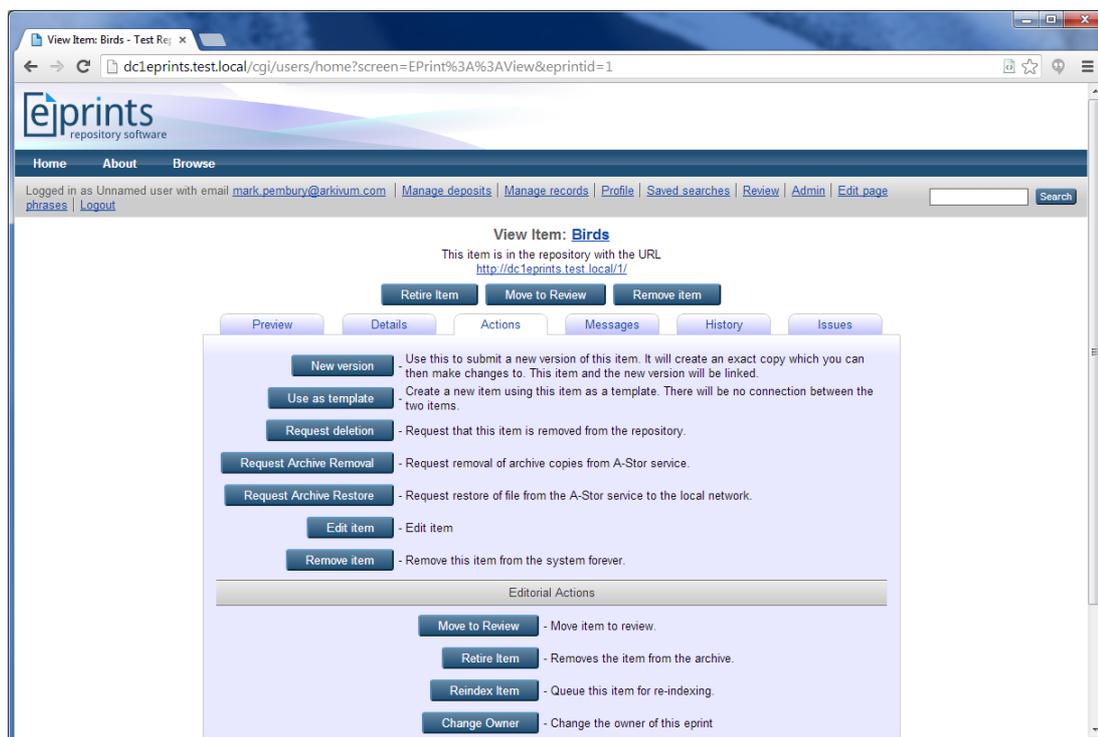
Once an item has been archived (escrowed) in the Arkivum service then the local copy of the item stored in EPrint will be removed. Only the copies stored in the Arkivum service will remain.

One further note on the archiving process is that an item will not be considered archived (or escrowed) until it has been written to escrow media and verified. The Arkivum service will write escrow media depending on the size threshold set-up for the data pool. E.g. If you set-up a data pool for 1.2TB then once 1.2TB has been stored an escrow set will be generated. You can also trigger a manual escrow set by logging into the A-Stor WebUI.

Only when this is done in the Arkivum service will the item then be marked as archived (escrowed) and the process completed.

5.2 Additional EPrint Actions

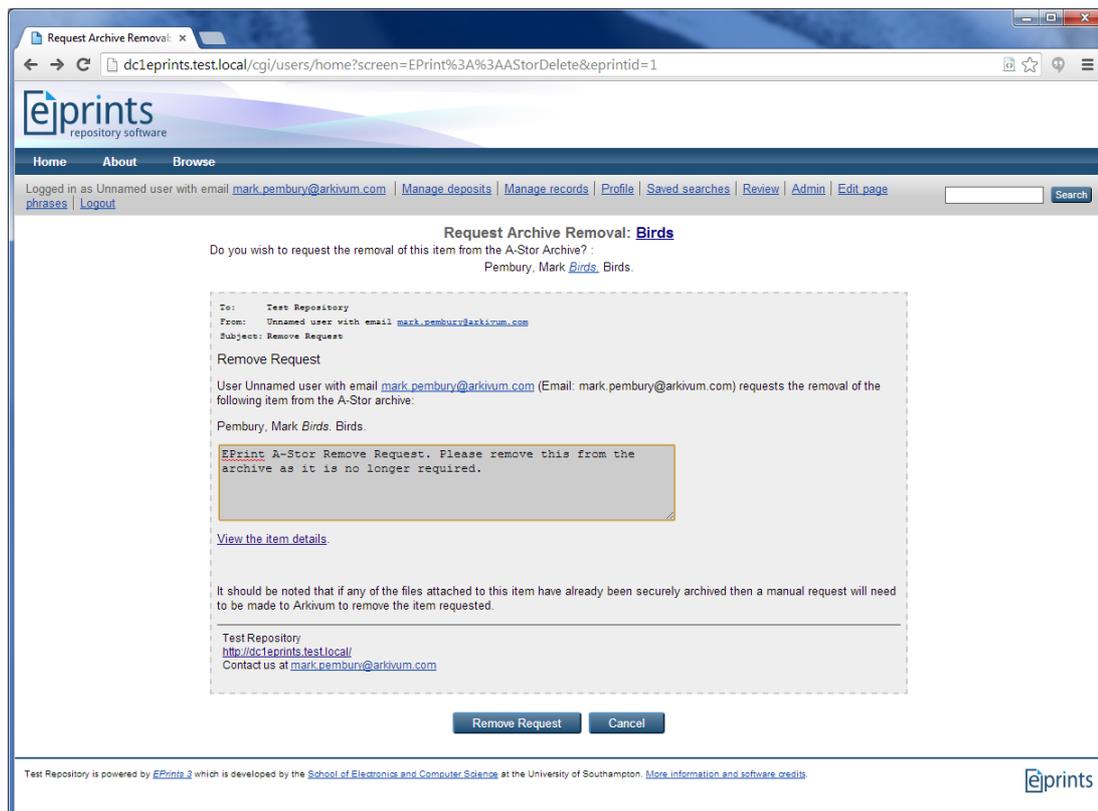
The version 2.0 plugin has added two additional plugins to the EPrint actions screen, Request Archive Removal and Request Archive Restore.



The Actions tab of an EPrint Item showing the two new action buttons.

5.3 Request Archive Removal

Once an EPrint item has been archive then an additional action called Request Archive Removal will become available on the actions tab of the EPrint item. This new screen will allow the user to request that the EPrint is removed from the A-Stor archive. If the request is approved then the EPrint will be removed from the A-Stor service.

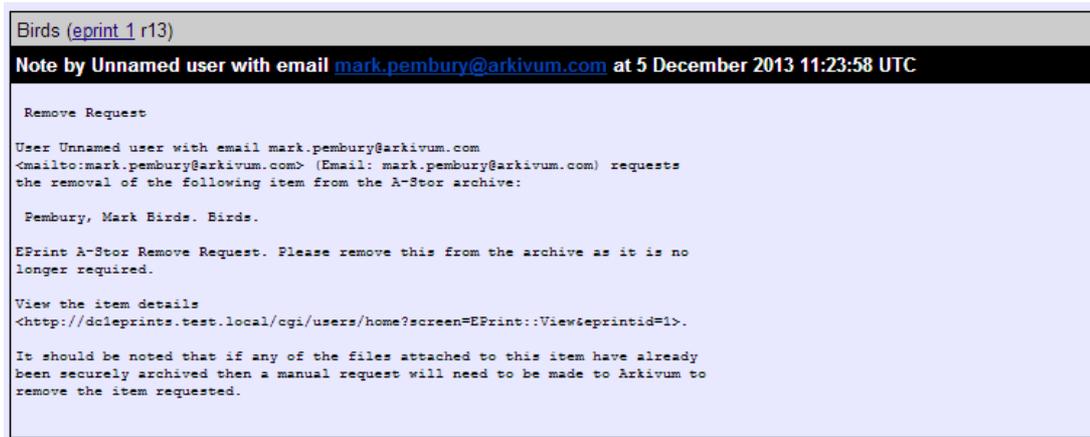


The Request Archive Removal Screen

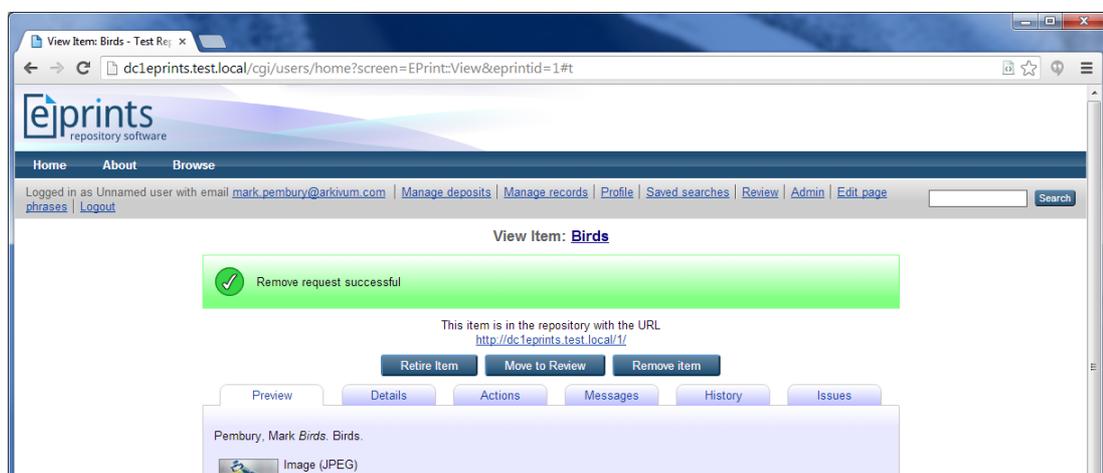
The screen has a link that will display a text box, allowing the user to enter a reason for the removal as well as a link so to the item itself.

Once the request is made a note is added to the item and the user will be returned to the item preview screen. The note can be viewed from either the messages or the history tab.

It should be noted that the removal itself make take some time. This will depend on the stage it has reaching in archiving in the A-Stor service. E.g. If the data has been replicated but not escrowed (archived) then it can be removed immediately. If it has been escrowed then this will require an action by Arkivum.



The note added to the EPrint item when a removal request is made



The preview pane after a removal request has been made

5.4 Request Archive Restore

Once an EPrint item has been archive then an additional action called Request Archive Restore will become available on the actions tab of the EPrint item.

This new screen will allow the user to request that the EPrint is restored from the A-Stor service to the local A-Stor appliance. If the request is approved then the EPrint will be scheduled for restore and the user notified by email when this has been completed. The item will also have a noted added to indicate that the restore has completed.

When an item is archived in the A-Stor service a copy of the item is stored on the local A-Stor appliance as well as in the data centers and on secure escrow media. Whenever EPrints requires a file that belongs to the EPrint item it will access that file via the mounted file system from the A-Stor appliance.

Over time the A-Stor appliance may need the space occupied by the appliance copy of the item and so will remove it. The appliance will restore this copy automatically from the data center when required, for example if a read request for the item is made from the mounted file system.

It should be noted that for the most part a user will not be required to restore an item whenever they wish to access it. This restore request will only be required if the following conditions are met:

1. The item is archived (escrowed) in the Arkivum service.
2. One or more file attached to the item are not stored on the local A-Stor appliance.

The request to restore a file can be made either through the action button or by the user accessing a file attached to the EPrint item and the file meeting the conditions stated above.

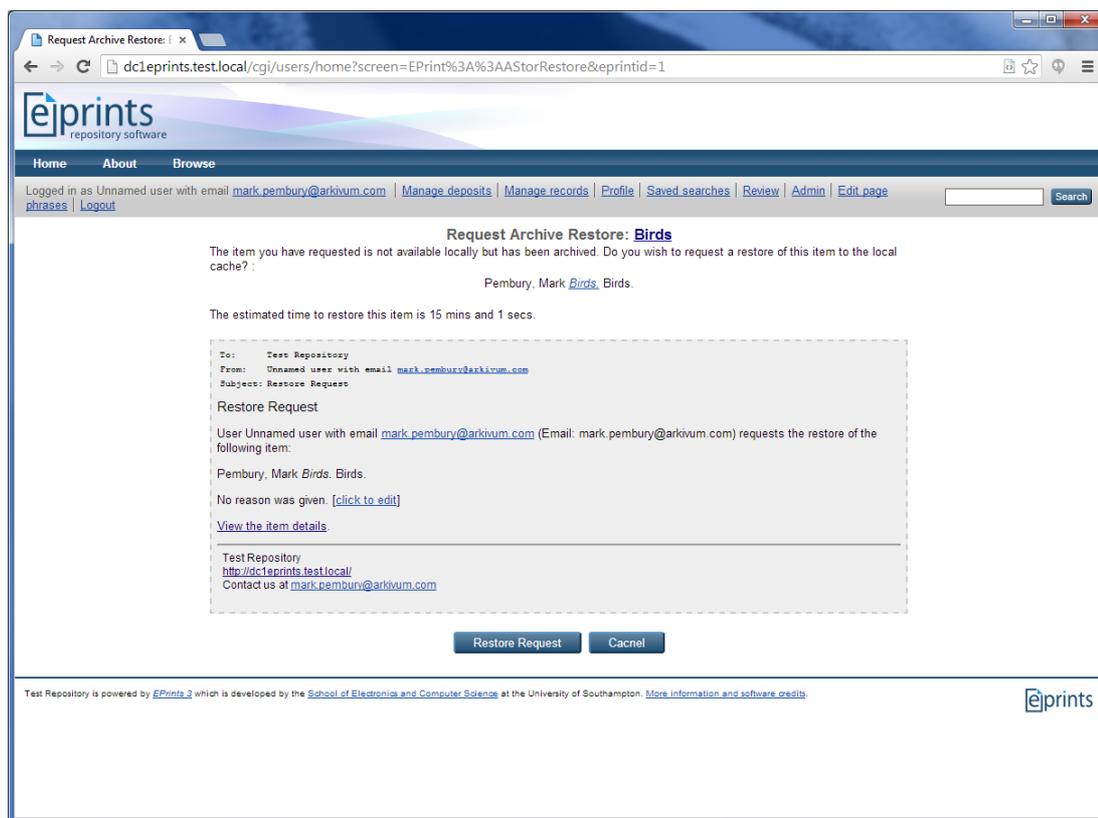
There is also an additional plugin configuration setting called size threshold. This value will determine if the file needs to be requested for a restore or can be restored automatically from the Arkivum service.

For example, if we have an EPrint item that has been successfully archived in the Arkivum service. The item has five pictures attached to it. Three of the pictures have a file size less than 1GB while the other two are over 2GB in size each.

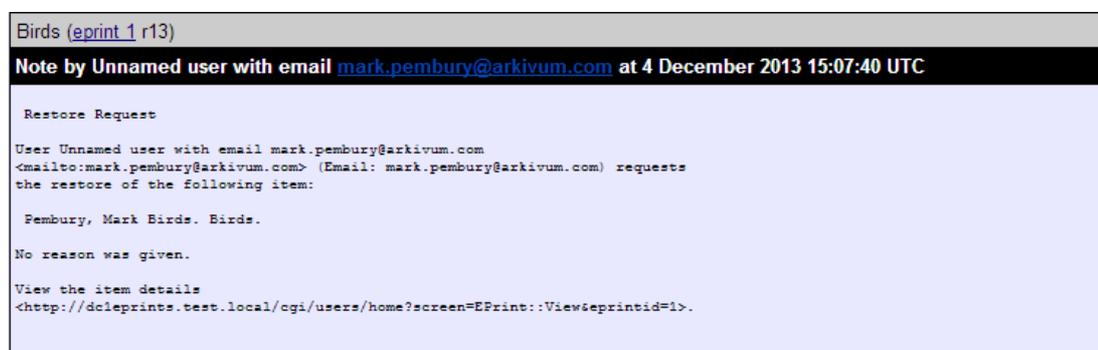
If a user click on a link to one of these pictures and the file was available on the A-Stor appliance it would be read immediately via the mounted file system.

If the user clicked on a link to one of the three pictures less than 1GB in size and it was not available on the local A-Stor appliance then it would be read immediately via the mounted file system but also restored the A-Stor appliance.

If the user clicked on a link for one of the two file over 2GB in size and it was not available on the local A-Stor appliance then the user will be redirected to the Request Archive Restore screen so a request to restore all items for the EPrint could be made.



The Request Archive Restore Screen



The restore request note added to the EPrint item

Mon 25/11/2013 12:50
Test Repository
EPrint Restore Notification

To: Mark Pembury

Retention Policy Deleted Items (30 days) Expires 25/12/2013

i This item will expire in 20 days. To keep this item longer apply a different Retention Policy.

EPrint Restore Notification

Locations Pembury, Mark [Locations](#). Locations. (Live Archive)

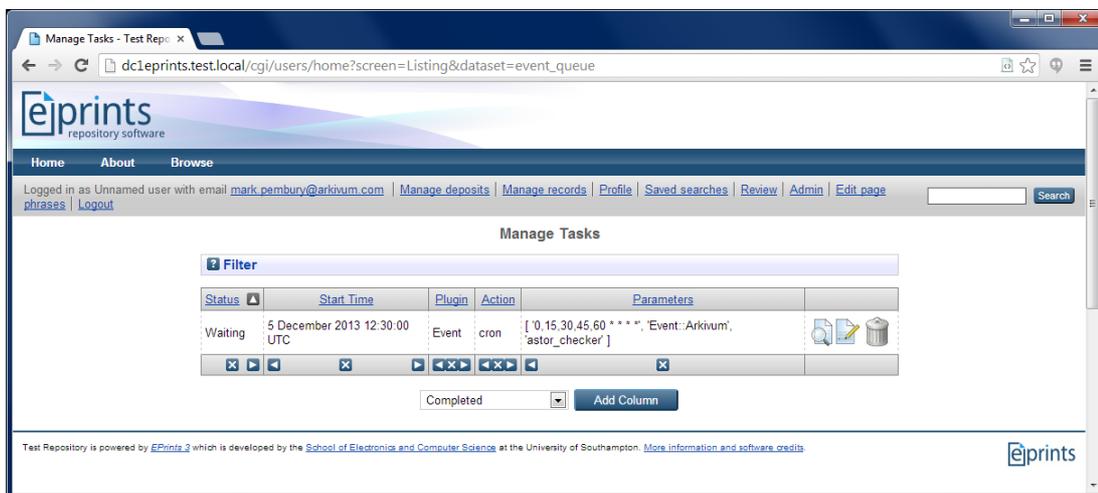
Has been restored and is now available.

Test Repository
<http://fluffy.arkivum.local/>
Contact us at mark.allsoff@arkivum.com

The email notification send to the restore requestor

5.5 A-Stor Checker Task

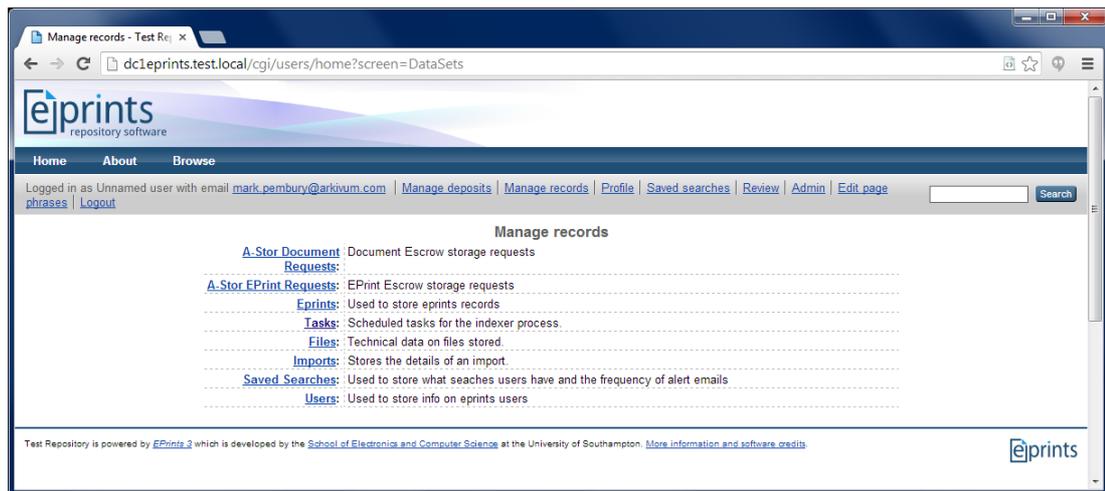
When the plugin is enabled it will add an indexer task that will check for requests to perform an action with the Arkivum service. By default this task will run every 15 minutes but this can be changed by configuration.



The A-Stor checker task displayed in the Manage Tasks screen in EPrints

5.6 Manage Records Screen

When the plugin is enabled it will add two additional datasets to the EPrints database. These two datasets manage the requests for either an EPrint item or a document item. These requests can be viewed and managed via the manage records screen.



The Manage Records screen with the two additional options.

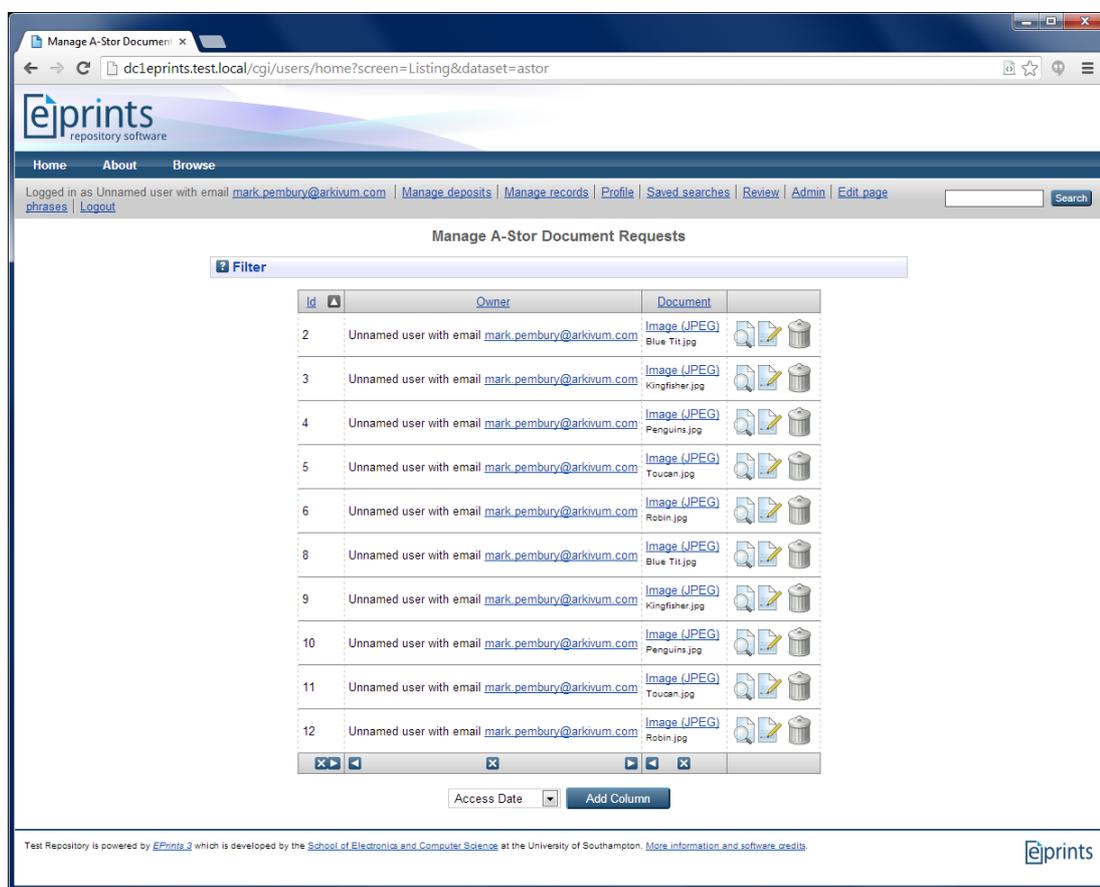
5.7 Manage Document Requests

This screen will allow an administrator to manage the requests for action on documents between EPrints and A-Stor.

For version 2.0 of the plugin no document authorization are required as they are managed automatically. Only authorization on EPrint requests may be required.

For version 1.0 this screen is used to authorize any request to archive a document.

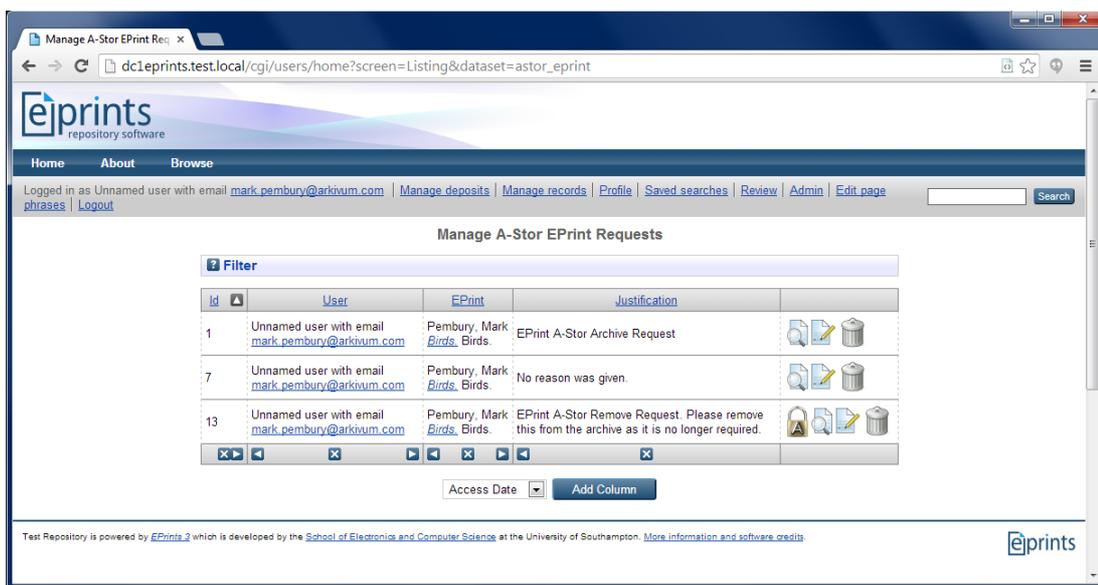
This screen can also be used to view requests that have failed.



The Manage A-Stor Document Requests screen

5.8 Manage EPrint Requests

This screen will allow an administrator to manage the requests to either restore or remove an EPrint from the Arkivum service. It will also show the status of an archive request automatically submitted when an item is moved to the repository.

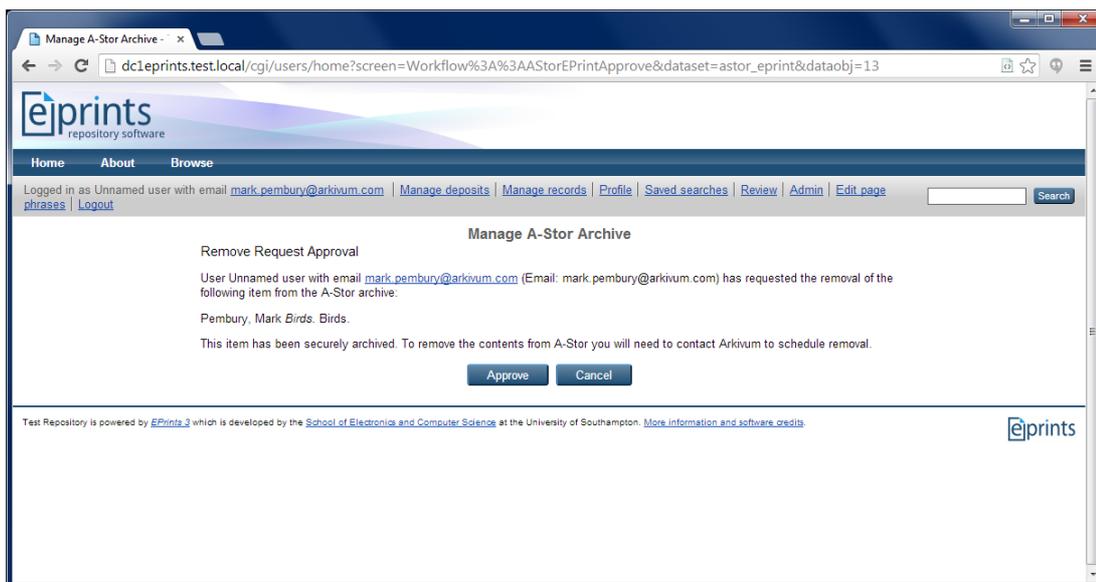


The Manage A-Stor EPrint Requests screen

In the screen shot above there is one item that requires authorization before the request will be scheduled.

5.9 Authorizing a Request

When a request is made that requires authorization it will be displayed in either the Manage A-Stor EPrint requests or Manage A-Stor Document Requests screen with the Arkivum icon next to it. Clicking on this icon will display the appropriate authorization screen.



A request to remove an EPrint item from the Arkivum Service

When a request to either restore or remove an EPrint item is approved a note is added to the EPrint item which can be displayed in the messages or history tab of the EPrint item.