

# NetarchiveSuite 5.5 Release Notes

- [Patch Fix "UmbraHookPatch"](#)
- [Highlights in 5.5](#)
- [Upgrading From Previous NetarchiveSuite Releases](#)
- [Issues Resolved in Release 5.5](#)

5.5 Release Date: 2018-12-04

## Patch Fix "UmbraHookPatch"

This patch enables one to execute a script to cleanup an umbra instance before starting a new harvest on any umbra enabled harvester. The patch can be enabled by replacing the two (identical) jar-files `heritrix3-controller-5.5.jar` and `netarchivesuite-heritrix3-controller.jar` with the jarfile [heritrix3-controller-UmbraHookPatch.jar](#) and restarting the HarvestControllerApplication instance. The git-source for this path is [commit acf5880a](#).

The settings for umbra in HarvestControllerApplication have an extra optional element

```
<umbra>
.
.
.
  <startupHook>drain-queue</startupHook>
</umbra>
```

The default value is "drain-queue" but this can be replaced with the path to a more sophisticated script - for example one that also restarts umbra. In tests we have used the following script to enable the specific python version under which umbra runs in the Netarkivet installation

```
#!/usr/bin/env bash
ulimit -c 0

source /opt/rh/rh-python36/enable

drain-queue
```

### Summary of installation steps

- Replace both jar files with the patched jar
- Create the script you wish to use and make it executable
- Modify the settings to point to the script
- Restart the HarvestControllerApplication

### Side effects

- All script output is logged in the HarvestControllerApplication log
- Remember that the default implementation "drain-queue" will empty the entire umbra input queue. So therefore it is highly inadvisable to have more than one HarvestControllerApplication using the same umbra instance
- The call to execute the hook script is blocking, so heritrix will not start until the script ends

## Highlights in 5.5

- NetarchiveSuite now supports browser-based harvesting using Internet Archive Umbra
- Improved stability in Heritrix MatchesRegexListDecideRule
- Improved handling of queue-assignments in Heritrix

## Upgrading From Previous NetarchiveSuite Releases

There are no special requirements involved in the upgrade. It should be sufficient to replace all .jar files in your installation lib directory with those from the new release, and replace the heritrix bundler zip-file on your HarvestController machines with the new bundler.

### Most-recent updates for 5.5:

- Download [NetarchiveSuite](#)
- Download [Heritrix 3 Bundle](#) (required)
- [Javadoc](#)
- [Manuals](#)

Enabling Umbra integration requires some reconfiguration. This is described in the [documentation](#). Note that if enabling Umbra, you should define the new queue for Umbra jobs in the NetarchiveSuite GUI *before* you start any new HarvestController instances to listen to the queue. (See

[NAS-2794](#) - Getting issue details...  .)

## Issues Resolved in Release 5.5

Key	Summary	Status
<a href="#">NAS-2797</a>	<a href="#">SeedUriDomainnameQueueAssignmentPolicy</a>	<input type="button" value="CLOSED"/>
<a href="#">NAS-2803</a>	<a href="#">Heritrix can hang on a pathological regex</a>	<input type="button" value="CLOSED"/>
<a href="#">NAS-2703</a>	<a href="#">Patch H3 to support srcset attribute</a>	<input type="button" value="CLOSED"/>
<a href="#">NAS-2514</a>	<a href="#">-50 when using the SeedUriDomainnameQueueAssignmentPolicy</a>	<input type="button" value="RESOLVED"/>
<a href="#">NAS-2522</a>	<a href="#">Check packaging of IA/NAS wayback war</a>	<input type="button" value="CLOSED"/>
<a href="#">NAS-2764</a>	<a href="#">Create a consistent Heritrix build for use with Umbra &amp; non-Umbra harvests</a>	<input type="button" value="CLOSED"/>
<a href="#">NAS-2792</a>	<a href="#">Fix H3 PAUSING status and add EMPTY status on Running job page</a>	<input type="button" value="CLOSED"/>
<a href="#">NAS-2804</a>	<a href="#">Build and Release Heritrix and Update Snapshot Dependency in NAS</a>	<input type="button" value="CLOSED"/>

[8 issues](#)