

NetarchiveSuite 5.0 Release Notes

Release Date: 17th September 2015

Contents

- Highlights
 - Heritrix 3
 - Java 8
 - More detailed jar file structure
 - Heritrix3 bundler zip now needs to be supplied in the deploy
 - Switch to Maven as build tool
 - Source moved to Github
- Download NetarchiveSuite
- Download Heritrix 3 Bundle (required)
- Javadoc
- Manuals
- Full list of issues resolved in this release
- Known issues

- **Download NetarchiveSuite**
- **Download Heritrix 3 Bundle (required)**
- **Javadoc**
- **Manuals**

Highlights

Heritrix 3

This is the first production release for NetarchiveSuite to support harvesting with Heritrix 3. Harvesting with Heritrix 1 is not supported in this version but may be supported in later versions.

Java 8

NetarchiveSuite now requires java 8 to run.

More detailed jar file structure

The jar files now fit the application structure better. This means the deploy script should be updated to the follow application <-> jar file classpath definitions. See [Updating deploy file jar definitions to 5.0](#).

Heritrix3 bundler zip now needs to be supplied in the deploy

The Heritrix3 crawler code has been moved into a separate zip file to allow more flexibility for choosing a concrete H3 version to use. This means that the HarvesterControllers deployment now needs to be configured with a extra H3 bundler zip. This can either be done as an [argument to the deploy call](#) or by defining a bundler for each [HarvestControllerApplication](#) in the [deploy configuration](#).

Switch to Maven as build tool

The project is now build with [Maven](#). This means that the jars, source-jars and javadoc jar can be found at the <https://sbforge.org/nexus/content/groups/public/> repository and also that netarchivesuite modules can be added as maven dependencies in other projects, for example:

```
<dependency>
  <groupId>org.netarchivesuite</groupId>
  <artifactId>nas-module</artifactId>
  <version>${nas.version}</version>
</dependency>
```

Source moved to Github

The Netarchivesuite source code is now located at github here: <https://github.com/netarchivesuite/netarchivesuite>.

Full list of issues resolved in this release

| T | Key | P | Summary |
|---|----------|---|--|
| | NAS-1780 | | Upgrade to or support for heritrix 3 |
| | NAS-2394 | | Replace CDX index rollover naming strategy |
| | NAS-2416 | | Bug in job generation. domain specific crawlertraps are not included in the template |
| | NAS-2434 | | Unable to access the H3 GUI in quickstart |
| | NAS-2436 | | vbscript fails when JAVA_HOME not set |
| | NAS-2454 | | Logging contaminates the output from our tools |
| | NAS-2333 | | migrate SeedUriDomainnameQueueAssignmentPolicy to H3 |
| | NAS-2376 | | Security manager should only be activated for Bitarchive application |
| | NAS-2397 | | harvesting-server should not get Invalid channelname, if the channel-name is unknown in the channels table |
| | NAS-2403 | | Deployment of H3 HC |
| | NAS-2410 | | Include the Icelandic H3 extractorJs in our heritrix3-extensions |
| | NAS-2418 | | Adding NAS CrawlMetadata with date field included |
| | NAS-2433 | | derbyrun-10.10.1.1.jar is not part of the libraries in NAS-5.0 |
| | NAS-2446 | | Review NetarchiveSuite Overview for NAS 5.0 |
| | NAS-2447 | | Review Quick Start Manual for NAS 5.0 |
| | NAS-2448 | | Review User Manual for NAS 5.0 |
| | NAS-2449 | | Review Installation Manual for NAS 5.0 |
| | NAS-2450 | | Review Configuration Manual for NAS 5.0 |
| | NAS-2451 | | Review System Design Manual for NAS 5.0 |
| | NAS-2452 | | Review Additional Tools Manual for NAS 5.0 |

Showing 20 out of 21 issues

Known issues

| T | Key | P | Summary | Fix Version/s |
|---|----------|---|---|---------------|
| | NAS-2474 | | Postgresql breaks connections to client prematurely | 5.1 |
| | NAS-2456 | | Error when adding a lot of seeds to harvestdefinition (jetty) | 5.1 |
| | NAS-2492 | | Error restarting jobs | 5.1 |
| | NAS-2467 | | Inefficient coding in BitarchiveAdmin | 5.1 |
| | NAS-2478 | | When the constructor of FTPRemote fails to upload the file to FTP-server, we don't get the reason | 5.1 |
| | NAS-2477 | | Heritrix3-controller should verify that the job has launched properly | 5.1 |
| | NAS-2419 | | GlobalCrawlertraps feature fails to read the whole file when creating a traplist (jetty) | 5.1 |

| | | |
|----------|---|-------|
| NAS-2596 | H3 pauseAtStart bean property ignored by Harvest Controller | 5.3 |
| NAS-2496 | Redirect for jp.dk fails in test wayback | 5.3 |
| NAS-2539 | NetarchiveSuite does not validate that domain crawlertraps are Wellformed XML | 5.5.1 |
| NAS-2420 | Need to XML escape Global Crawlertraps | 5.5.1 |
| NAS-2430 | Log for HarvestController no longer provides link to Harvester webGUI | |

12 issues