



netarchive.dk

---

# Settings in the NetarchiveSuite

Configuring the properties of the  
NetarchiveSuite and its components  
Søren V. Carlsen (svc@kb.dk)



## Settings and Configurations

---

- ***Configuring the netarchiveSuite is done using***
  - *Settings files (default location `./conf/settings.xml`. The location of the settings file is normally overridden on the command line using - `Ddk.netarkivet.settings.file=/path/to/the/correct/settings.xml`) ↴*
  - *Miscellaneous configuration done on the commandline when starting applications*



## Settings files

---

- *The setting files contain 4 main section, each related to one of the 4 main java packages in NetarchiveSuite:*
    - *common (dk.netarkivet.common),*
    - *harvester (dk.netarkivet.harvester),*
    - *archive (dk.netarkivet.archive),*
    - *viewerproxy (dk.netarkivet.viewerproxy) ↵*
    - *monitor (dk.netarkivet.monitor)*
    - *"deploy" ↵ (dk.netarkivet.deploy) "*
  - *Link to [settings\\_example.xml](#) (from NetarchiveSuite 3.2.2) ↵*
-



## Common Settings (1)

---

- *The common settings contain settings for the:*
    - *Location of tempDir*  
*(settings.common.tempDir) ↵*
    - *RemoteFile plugin*  
*(settings.common.remoteFile.\*) ↵*
    - *JMS plugin*  
*(settings.common.jms.\*) ↵*
    - *Notification plugin*  
*(settings.common.notifications.\*)* ↵
-



## Common Settings (2)

---

- *And also contain settings for the:*
  - *arcrepositoryClient plugin (Archive access client)*
    - *settings.common.arcrepositoryClient.\**
  - *indexClient plugin (Client to request and receive indices)*
    - *settings.common.indexClient.\**
  - *JMX setup used by the application*
    - *settings.common.jmx.port*
    - *settings.common.jmx.rmiPort*
    - *settings.common.jmx.passwordFile*



## Common Settings (3)

---

- *And finally contain settings for related to the GUI:*
  - *Languages*
    - *settings.common.webinterface.language.locale*
    - *settings.common.webinterface.language.name*
  - *Sitesection plugin*
    - *Settings.common.webinterface.sitesection.\**

*All plugins will be described in depth later in the "Plugins" talk*

---



## Harvester Settings (1)

---

- *The harvester settings consist of settings related to*
  - *Domains and DomainConfigurations*  
*(settings.harvester.datamodel.domain.\*)*
    - *defaultSeedlist - Default seedlist*
    - *defaultConfig - Default domainconfiguration*
    - *defaultOrderxml - Default harvest template*
    - *validDomainRegex - Regular expression for valid domains*



## Harvester Settings (2)

---

- *The harvester settings consist of settings related to*
  - *Database setup*  
*(settings.harvester.datamodel.database.\*)*
    - *url (Database connection URL)*
    - *specificsclass (plugin)*
    - *backupInitHour (backup start time)*





## Harvester Settings (3)

---

- ❑ The settings for domains and domainconfiguration and Database setup - only used by the applications accessing the database (the GUI, scheduler, HarvestTemplateApplication).
  - ❑ Settings for jobcreation and scheduling - not discussed (*settings.harvester.scheduler.\**)
  - ❑ Settings for harvesting - used solely by the harvesters (HarvestControllerApplication) (*settings.harvester.harvesting.\**)
    - ❑ *serverDir* (dir for creating job directories)
    - ❑ *oldjobsDir* (dir where old jobs are moved)
    - ❑ *queuePriority* (From Which job pool should the HarvestController take its jobs. This value is either *LOWPRIORITY* (The pool used for snapshot harvests), or *HIGHPRIORITY* (The pool used for selective harvests))
    - ❑ *heritrix.inactivityTimeout* (when to terminate embedded Heritrix due to inactivity) - uses *crawlController.activeToeCount()* to measure inactivity
    - ❑ *heritrix.noresponseTimeout* (when to terminate embedded Heritrix on account of no data received from webservers)
-



# Archive settings (1)

---

The 'archive' section have settings relevant for the

- `arcrepository` application (`settings.archive.arcrepository.*`)
  - `baseDir` – The directory where the `admin.data` file is found. The `admin.data` file contains a list of all files stored in the archive, and their checksums
  - `location` (the name of all bit archive locations)
  - `batchLocation` (the default bitarchive to use for batch jobs)
- `BitarchiveServer` applications (`settings.archive.bitarchive.*`)
  - `fileDir` (directory to store ARCFiles. Each server can have more than one `fileDir`)
  - `minSpaceLeft` (minimum amount of free disk space required to accept ARC-files)
  - `heartbeatFrequency` (how often the server must send an `HeartBeatMessages` to its `BitarchiveMonitor`)
  - `limitForRecordDatatransferInFile` (If this limit is passed, the data is transferred using a `RemoteFile`, otherwise the data is transferred embedded in a JMS message)



## Archive settings (2)

---

The 'archive' section have settings relevant for

- the BitarchiveMonitor applications (*settings.archive.bitarchive.\**)
  - acceptableHeartbeatDelay (Maximum acceptable delay between heartbeats from a BitarchiveServer)
  - batchSizeTimeout (Maximum time to wait for a Batch job to finish - default 14 days)

Settings common for all three archive applications  
(*settings.archive.bitarchive.\**)

- thisLocation (Which location do this application belong to)
- credentials (password to be able correct/delete arcfiles)



## Viewerproxy and monitor settings

---

The 'viewerproxy' section have settings relevant for initializing and accessing the Viewerproxy application  
(`settings.viewerproxy.*`)

- `hostName` (The name of the server used for viewerproxy links in the GUI)
- `baseDir` (The main directory for the ViewerProxy Application)

The 'monitor' section have settings relevant for monitoring the NetarchiveSuite(`settings.monitor.*`)

- `applicationName` - The name of the application, fx. "BitarchiveServerApplication".The monitor puts this with each log message
- `logging.historySize` (# of logmessages from each application visible from the monitor part of the GUI)



## Configuration on the commandline (1)

---

Most settings in the settings files can be overridden on the command line using `-Dsettings.common` (e.g.

```
-Dsettings.common.tempDir=/Mytmpdir)
```

However this is not the case with settings that have one more

than a single values, like

- Languages  
(`settings.common.webinterface.language.*`)
- Sitesections  
(`settings.common.webinterface.siteSection.*`)
- `fileDir` (*`settings.archive.bitarchive.fileDir`*)



## Configuration on the commandline (2)

---

Some configuration of NetarchiveSuite can only be done on the command line as arguments for the applications. This is the case for

- Configuration of logging
  - **Pointing to the correct logger class:**
    - `-Dorg.apache.commons.logging.Log=org.apache.commons.logging.impl.Jdk14Logger`
  - **Pointing to the correct logging properties file**
    - `-Djava.util.logging.config.file=/home/test/SVC/conf/log.prop`
- Pointing to a settings file different from the default.  
E.g.
  - `”-Ddk.netarkivet.settings.file=/home/test/SVC/conf/my_settings.xml`