

Bit Preservation

- [Missing Files](#)
- [Checksum Errors](#)
- [Batchjob Overview](#)
- [Batchjob Checksum](#)
- [Batchjob Filelist](#)

The **Bitpreservation** interface lets you control active checks of the status of the underlying bitarchive. This only applies if your installation uses the NetarchiveSuite archive repository component.

The screenshot displays the Bit Preservation interface. On the left is a 'Menu' sidebar with a red binary logo (10101 101 1 101 10101) and the following items: Definitions, Harvest status, Harvest Channels, Bitpreservation (with sub-items Filestatus and Batchjob Overview), Quality Assurance, and Systemstate. The main content area is titled 'Filestatus' and shows the status for three bitarchive instances: CSN, SBN, and KBN. For each instance, it lists the number of files (16), missing files (0), and the last update time (all are '-'). Below each instance's status are links to 'Update filestatus' and 'Update checksum and filestatus'. The CSN instance is a checksum replica, while SBN and KBN are datafile replicas.

Filestatus

Status of the replicas

Filestatus for: **CSN**
Number of files: 16
Missing files: 0
Last update at -
[Update filestatus for CS](#)

Filestatus for: **SBN**
Number of files: 16
Missing files: 0
Last update at -
[Update filestatus for SB](#)

Filestatus for: **KBN**
Number of files: 16
Missing files: 0
Last update at -
[Update filestatus for KB](#)

Checksum status for: **CSN**
Number of files with error: 0
Last update at -
[Update checksum and filestatus for CS](#)

Checksum status for: **SBN**
Number of files with error: 0
Last update at -
[Update checksum and filestatus for SB](#)

Checksum status for: **KBN**
Number of files with error: 0
Last update at -
[Update checksum and filestatus for KB](#)

The interfaces lets you initiate two types of checks on every copy of the files in the archive: Filestatus and Checksum status.

In the example on the screen dump there are three bitarchive instances – *CSN*, *SBN* and *KBN*. *CSN* is a checksum replica - it contains only checksums, not copies of the datafiles.

The **Update** buttons let you update the status for both files and checksums for both of the full bitarchive instances. The page will give you the Filestatus as

- Number of files
 - Missing files
 - Last updated at and the Checksum status as

 - Number of files with error
 - Last updated at
- The Filestatus checks are rather fast because only the existence of the files are checked whereas the Checksum status checks can take days /weeks for larger archives depending on the number of CPUs and the IO-speed of your hard drives.

Missing Files

 **Menu**
[Definitions](#)
[Harvest status](#)
[Harvest Channels](#)
[Bitpreservation](#)
 [Filestatus](#)
 [Batchjob Overview](#)
[Quality Assurance](#)
[Systemstate](#)

Missing Files

Missing files for SBN:

Status:

| | | |
|--|--------------------------|----------|
| 13-4-20161107235039550-00000-kb-test-har-003.kb.dk.warc.gz | <input type="checkbox"/> | Get info |
| 14-metadata-1.warc.gz | <input type="checkbox"/> | Get info |

Change the infobox for files

If files are missing on one instance of the bitarchive a "Show missing files" link will appear right next to the line with the number of missing files

For each missing file you can select "Get info". With the "Change the infobox for"-field in the bottom of the screen you can select a number of files in one operation.

Pressing "Execute" makes the system get a fresh status on the files and their checksums from both copies (out of which one is missing) and from the administrative system so that there are always three checksums available for each file.



Menu

- Definitions
- Harvest status
- Harvest Channels
- Bitpreservation
 - Filestatus
 - Batchjob Overview
- Quality Assurance
- Systemstate

Missing Files

Missing files for SBN:

Status:

13-4-20161107235039550-00000-kb-test-har-003.kb.dk.warc.gz Get info

| Replica | Admin State | Checksum | |
|------------|------------------|----------------------------------|---|
| Admin Data | - | NO ADMIN CHECKSUM! | |
| CSN | UPLOAD_COMPLETED | b73489ee5ab2367c06b16e97d7d85e89 | Add to replica SBN <input type="checkbox"/> |
| SBN | UPLOAD_FAILED | No checksum | |
| KBN | UPLOAD_COMPLETED | b73489ee5ab2367c06b16e97d7d85e89 | |

14-metadata-1.warc.gz Get info

| Replica | Admin State | Checksum | |
|------------|------------------|----------------------------------|---|
| Admin Data | - | NO ADMIN CHECKSUM! | |
| CSN | UPLOAD_COMPLETED | d3fcdcc75bab64b2f370f2a7b4729e43 | Add to replica SBN <input type="checkbox"/> |
| SBN | UPLOAD_FAILED | No checksum | |
| KBN | UPLOAD_COMPLETED | d3fcdcc75bab64b2f370f2a7b4729e43 | |

Execute

- Change the infobox for files
- Change that can be added

If the two "remaining" checksums (On the screen dump 32 byte long) are identical the system allows you to add the missing file to the bitarchive instance that had lost it.

This requires you to click "Add to archive" and then press "Execute". Marking files for addition can be done for multiple files at once.

Checksum Errors



Menu

- Definitions
- Harvest status
- Harvest Channels
- Bitpreservation
 - Filestatus
 - Batchjob Overview
- Quality Assurance
- Systemstate

Missing Files

The file 14-metadata-1.warc.gz has been restored in replica SBN

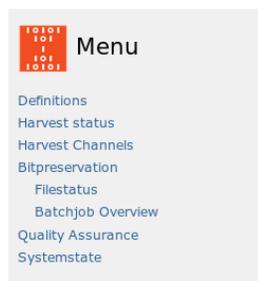
Missing files for SBN:

There are no more missing files

In case of a checksum error this error can be corrected through the interface.

To replace a bad file you need to type a security password and press **Replace the file in bitarchive replica XX**. The bad file will not be completely removed but moved to an 'attic' directory on the bitarchive server holding the bad file.

Batchjob Overview



Menu

- Definitions
- Harvest status
- Harvest Channels
- Bitpreservation
 - Filestatus
 - Batchjob Overview
- Quality Assurance
- Systemstate

Batchjob Overview

| Batchjob | Last run | Output file | Error file |
|-----------------------------|------------------------------|--|---|
| ChecksumJob | Tue Nov 08 14:06:44 CET 2016 | Download outputfile 149 bytes, 2 lines | Download errorfile 256 bytes, 4 lines |
| FileListJob | Batchjob has never been run | No outputfile | No errorfile |

To see the overview of batchjobs select “Batch overview” on the left menu.

Press the **ChecksumJob** link in the batchjob-column to get ready to run one or more checksum jobs.



Menu

- Definitions
- Harvest status
- Harvest Channels
- Bitpreservation
 - Filestatus
 - Batchjob Overview
- Quality Assurance
- Systemstate

Batchjob

Name of batchjob: `dk.netarkivet.common.utils.batch.ChecksumJob`

Number of runs: 1

| Started date | Ended date | Output file | Error file |
|------------------------------|------------------------------|---|--|
| Tue Nov 08 14:06:44 CET 2016 | Tue Nov 08 14:06:46 CET 2016 | 149 bytes 2 lines Download outputfile | 256 bytes 4 lines Download errorfile |

Choose replica:

- CSN CHECKSUM
- SBN BITARCHIVE
- KBN BITARCHIVE

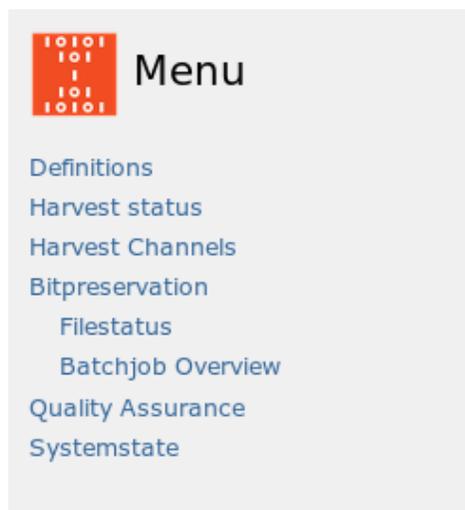
Which files:

Job ID:

- Metadata
- Content
- Both

Batchjob Checksum

You can select which replica to run the batchjob on and restrict the jobs on which the batchjob should run. Press the **Execute batchjob** button to start the desired checksum.



Menu

- Definitions
- Harvest status
- Harvest Channels
- Bitpreservation
 - Filestatus
 - Batchjob Overview
- Quality Assurance
- Systemstate

Executing batchjob

Executing batchjob with the following parameters.

BatchJob name: `dk.netarkivet.common.utils.batch.ChecksumJob`

Replica: SBN

Regular expression: `.*1-.*metadata.*`

When you press the link **Download outputfile**, the system starts a standard download dialog to present the .txt file e.g. in notepad

When you press the link **Download errorfile**, the system starts a standard download dialog to present the .txt file e.g. in notepad

Batchjob Filelist

Press the **FileListJob** link in the batchjob-column to get ready to run one or more filelist jobs.



Menu

- Definitions
- Harvest status
- Harvest Channels
- Bitpreservation
 - Filestatus
 - Batchjob Overview
- Quality Assurance
- Systemstate

Batchjob

Name of batchjob: **dk.netarkivet.common.utils.batch.FileListJob**

Batchjob has never been run

Choose replica:

- CSN CHECKSUM
- SBN BITARCHIVE
- KBN BITARCHIVE

Which files:

Job ID

- Metadata
- Content
- Both

Press the **Execute batchjob** button to start the desired Filelist

