

---

## Cleanup Job

- **Product:** Rosetta
  - **Product Version:** 5+
  - **Relevant for Installation Type:** Local
- 

### Description

All cleanup activities are logged `/operational_shared/logs/server.log`  
Directories that are actively used by the Rosetta application are not cleaned with the Cleanup Job.  
Where possible, technical, business, and performance cleanup is executed in real time.

To review, change, or schedule the Cleanup Job:

1. Connect to the Administrative Module and navigate to Scheduled Jobs Management
2. Click "Update" for the Cleanup Job

You can schedule the Cleanup Job to run hourly, daily, weekly, monthly, or as a cron (advanced).  
Deletes are made in the DB (HDpDepositActivity and sip\_registry) and file system.

Below are the list of Job Parameters that are inactive by default:

#### 1. Clean staff\_work\_area Directories

Default parameter: Older than days: 30

Source Directory: `/operational_shared/staff_work_area`

Notes: Cleans all files

#### 2. Clean Delivery Cache

Default parameter: Older than days: 90

Notes: Cleans files and thumbnail creation for delivery viewers (in database)

#### 3. Clean Update Metadata Job Directories

Default parameter: Older than days: 30

Source Directory: `/operational_shared/md_update`

Notes: Cleans files from 'done' subdirectory.

#### 4. Clean Add Derivative Rep Job Directories

Default parameter: Older than days: 30

Source Directory: `/operational_storage/derivative_copies`

Notes: Cleans only METS + files directories with 'DONE' file

#### 5. Clean Old Deposit Jobs

Default parameter: Older than days: 183

Source Directory: /deposit\_storage/1-1000

Notes: Clean files and empty parent folders as follows:

File system: Cleans directories and caches.

Database: Cleans records from HDpDepositActivity with statuses DECLINED or APPROVED, older than the input parameter (e.g. older than 30 days).

Also, regardless of "Clean Old Deposit Jobs" checkbox state in Cleanup Job screen, running Cleanup Job cleans from HDpDepositActivity all existing DELETED status records and also INCOMPLETE status records older than 7 days.

## **6. Clean Preservation Alternative Plans Directories**

Default parameter: Older than days: 90

Source Directory: /operational\_shared/operational\_export\_directory

Notes: Cleans the Preservation directories and files exported from several places within Rosetta

## **7. Clean Finished SIPs**

Default parameter: Older than days: 183

Source Directory: /operational\_shared/sipTmpDir

Notes: Used as a temporary directory for uploaded file when replacing files.

Clean files and empty parent folders as follows:

Database: Cleans all records from the sip\_registry DB table that have a "FINISHED" status and are older than the input parameter (e.g. older than 30 days).

SIPs are deleted only after its related deposit\_activity is deleted.

File system: Cleans only SIPs (XMLs + files) that are in the permanent repository.

## **8. Clean Deleted OAI-PMH Records**

Default parameter: Older than days: 90

Source Directory: /operational\_shared/<location varies depending on submission format definition>

Notes: It is recommended that you set the OAI deletedRecord policy to transient in the oaiproviderconfig.xml configuration file.

## **9. Clean Old Events**

Default parameter: Older than days: 365

All / Exclude / Include Event Types:

Notes: Can be configured to include / exclude specific event types in the audit event table based on date and event type.

Removes HFREVENT and HFREVENTKEYS.

## **10. Clean Process History**

Default parameter: Older than days: 183

Source Directory: /operational\_shared/sipTmpDir

Notes: Cleans any orphan process automatically (result of delete process)

## **Additional Information**

### **Other directories that are cleaned:**

/operational\_shared/backoffice/system\_jobs (cleans files used for system jobs)

/operational\_shared/bytestream\_work (used as a temporary directory for bytestream processing)

/operational\_shared/format\_library\_downloads (cleans old Format Library downloaded files)

/operational\_shared/tmp/metsTmpDir (cleans only XMLs and files / directories from Move to Permanent that are now in the permanent repository)

/operational\_shared/tmp/ScheduledReportsFiles (cleans BIRT Excel, Word, PDF, etc. report output).

/operational\_shared/tmp/Scripts (cleans files not in use by Rosetta)

/operational\_shared/tmp/storage/destPath (cleans small text files used by the NFS Plugin)

- 
- **Article last edited:** 26-Mar-2025