

14. Archive Management/Data Pool Maintenance

14.1 Archive Management Overview

Archive processing is at the core of the ECS system. The Online Archive consists of the following components:

- Data Pool/Data Archive: a persistent data store for all science and ancillary data,
- Browse Archive: a persistent data store for all browse data files,
- Small File Archive: a persistent data store for all ESDT definition files, and ESDT specific XML schema files, and a backup copy of all Science Granule XML Metadata files.

The access to the Data Pool/Data Archive is controlled, but all publicly available data holdings are accessible and directly downloadable by the users through the public Data Pool. The product orders for the non-public data holdings are first retrieved from the hidden area of the Data Pool/Archive and placed into the staging area and distributed to the users.

The disk-based archive is augmented with a backup tape archive system primarily for disaster recovery purpose. The tape archive holds a copy of the data found in the disk archive. Since all data resides on the disk-base archive, there is a smaller probability of having to recall data from tape, so the number of tape devices has been reduced from the previous tape near-line archive configuration.

14.2 Archive Hardware

The Archive hardware can be broken down into four groups:

1. StorNext Storage Manager (SNSM) Metadata Controller Servers.
2. SAN Fabric.
3. Disk Based Archive.
4. Managed Storage.

SNSM Metadata Controller Servers (MDC) are two identical HP ProLiant ML570 servers configured in a highly available (HA) configuration. The MDCs are using the Linux Red Hat Linux 5 Update 2 operating system. These servers host the software that manages the storage connected to the fabric in the ECS system architecture. The MDC Failover allows a secondary MDC host to take over StorNext operations in the event a primary MDC host fails. Failover is supported for all StorNext management operations, including client I/O requests (File System) and data mover operations (Storage Manager). This significantly increases the uptime of the ECS archive at each DAAC. The metadata server is the traffic cop that tracks the shared file

system activity and maintains the synchronization of updates. The MDCs are connected to the fabric and the private network to automatically move data between disk and tape archives.

The **SAN Fabric** consists of two Brocade 4100 switches connecting to all hosts and storage arrays with a view into the Data Pool. The switches allow attached hosts access to the Data Pool through the fabric. A private gigabit ethernet network carries the metadata communications between the hosts and the metadata server, while the fabric carries the data holdings.

The **Disk Based Archive** is EMC CLARiiON CX series disk arrays. The Data Pool, Browse, Smallfiles Archives along with the StorNext Archive cache are all located on this hardware. The Data Pool is spread across all controllers primarily using 1 TB drives bound as RAID 6. The Browse and Smallfiles Archives are on fibre channel in most locations to take advantage of the higher performance required for small writes. These disks are bound as RAID 5 which also has less overhead than RAID 6. The cache is a group of spindles have been set aside for StorNext to stage and de-stage data from media. The disks arrays are connected to the SAN Fabric for data transfers. The private network is used to manage the arrays using Navisphere.

The **SNMS Managed Storage** is connected to a separate tape SAN. The SNSM metadata servers and the tape drives are connected to the Tape Fabric. This SAN is used to move data from the StorNext Archive cache to LTO tape. A Scalar i6000 or i500 library have been installed at each DAAC with 6 to 8 LTO 4 tape drives connected to a storage networking blade. Each LTO 4 tape can store 800GB of data native (up to 1600 GB compressed). Slots are available for up to 128 tapes in the NSIDC and EDF libraries, 300 tapes in the PVC library, and 700 tapes in the ASDC and LP DAAC libraries. Each LTO tape cartridge is identified by a colored bar code label that shows the media number. An archive catalog or database tracks the location of each cartridge within the library, based on information provided by the laser bar code reader. Offline data storage racks are provided to store LTO tapes outside of the library. (Figure 14-2.1)

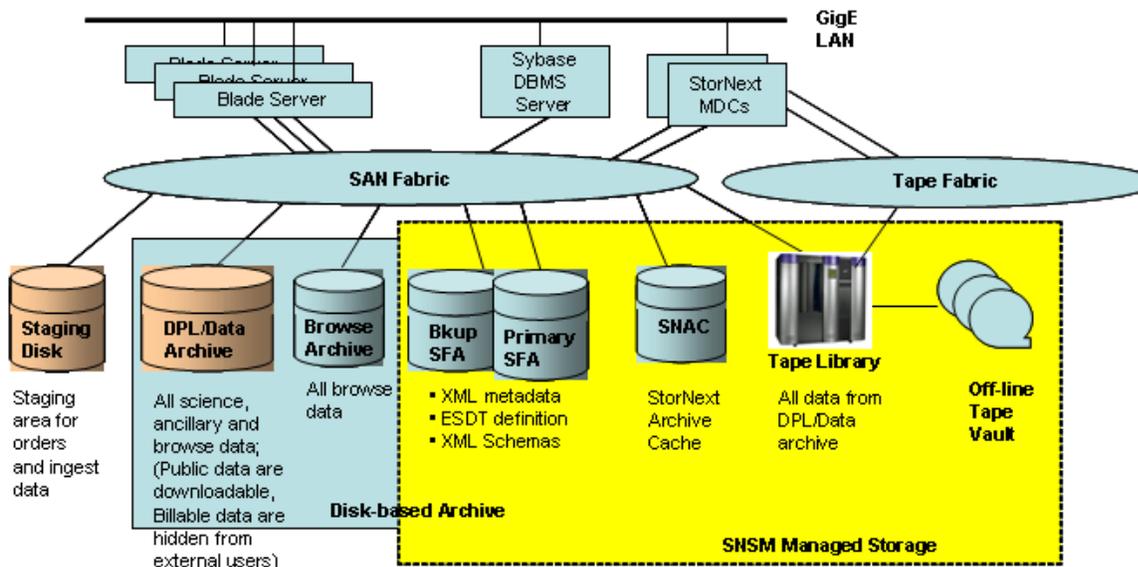


Figure 14.2-1. Online Archive Architecture

14.3 Archive Software

Archive operations rely on both custom and commercial off the shelf (COTS) software for complete mass storage archive management, providing the capability to accept Graphical User Interface (GUI) and command line interface inputs to control and monitor archive operations. The archive software is the Quantum’s StorNext Storage Manager (SNSM) software.

The StorNext Product has two components:

1. StorNext File System (SNFS) – is a shared file system called CVFS.
2. StorNext Storage Manager (SNSM) – this manages the volumes (media), drives and jukeboxes.

Primary dependencies are on the UNIX Index Node (inode) structures. StorNext maintains all inode information in database files rather than in associated disk structures. This minimizes or eliminates many of the file search problems inherent in searching large numbers of files in multiple directories.

14.4 Starting and Stopping StorNext

The ECS System contains both managed (Hierarchical Storage Manager) and unmanaged StorNext File Systems. In order for the ECS System to function properly you need to start/stop

both. They can be started/stopped from the Linux command line or from the GUI provided by the vendor. Table 14.4-1 provides an Activity Checklist for Starting and Stopping StorNext.

Table 14.4-1. Starting and Stopping StorNext

Order	Role	Task	Section
1	System Administrator or Archive Manager	Start the StorNext Manager Server (from the command line prompt)	(P) 14.4.1.1
2	System Administrator or Archive Manager	Start the StorNext manager Server (from the GUI)	(P) 14.4.1.2
3	System Administrator or Archive Manager	Stop the StorNext Clients	(P) 14.4.2.1
4	System Administrator or Archive Manager	Disabling the Tape Archive System	(P) 14.4.2.2
5	System Administrator or Archive Manager	Rebooting the StorNext Metadata Servers	(P) 14.4.3.1
6	System Administrator or Archive Manager	Creating a Listing of StorNext Labels	(P) 14.4.4.1

14.4.1 Starting the StorNext Application

To start the StorNext System, both the server and its clients must be started.

14.4.1.1 Start the StorNext Manager Server (from the command line prompt)

1 Logon to the active metadata server (x4smvaa) as root or superuser. Using x4smvaa, log into the active (x4sml01 - primary or x4sml02 - failover/secondary).

service cvfs start

2 Verify that the StorNext Manager Server is active with the command:

ps -ef | grep cvfs

To start StorNext Manager Clients from the command line prompt: (as superuser or root).

3 Logon to each of the clients.

service cvfs start

4 Verify that the StorNext Manager Server is active with the command:

ps -ef | grep cvfs

The **StorNext Home** GUI is Web based, and can be accessed by any current Web enabled machines with the proper Java libraries.

Note: Persons with Administrators Accounts can only have full control of the archive.

14.4.1.2 Start the StorNext Manager Server (from the GUI)

- 1 Open a Web browser. Mozilla is the EED supported standard, however, other browsers such as Firefox, Netscape 7+, and others may be used.
- 2 Enter the name of the active StorNext metadata server.

Example: **p4smvaa.pvc.ecs.nasa.gov:81**

- The **StorNext** login window will appear.
- 3 Enter the **username** (admin) and **password** in the spaces provided. (Operators can create multiple accounts.)
 - The **StorNext** GUI Home page will be displayed (Figure 14.4-1).

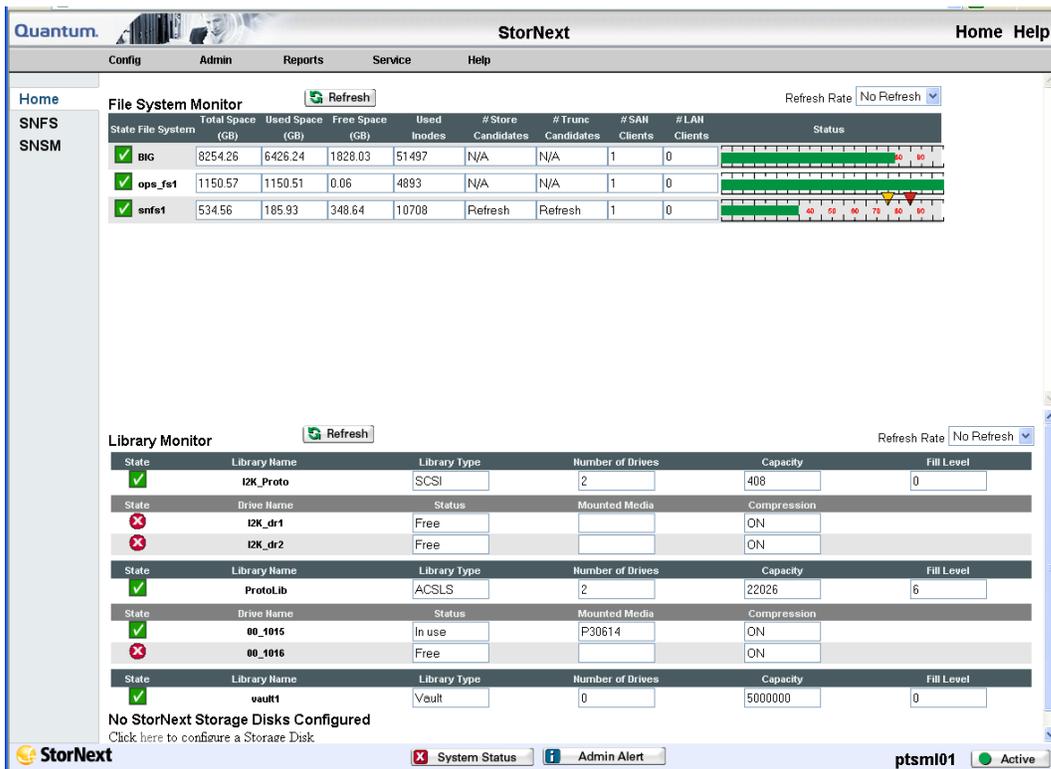


Figure 14.4-1. StorNext GUI Home Page

- 4 Select **Admin** from the Home Page.
 - The **Admin** pull-down menu will be displayed (Figure 14.4-2).

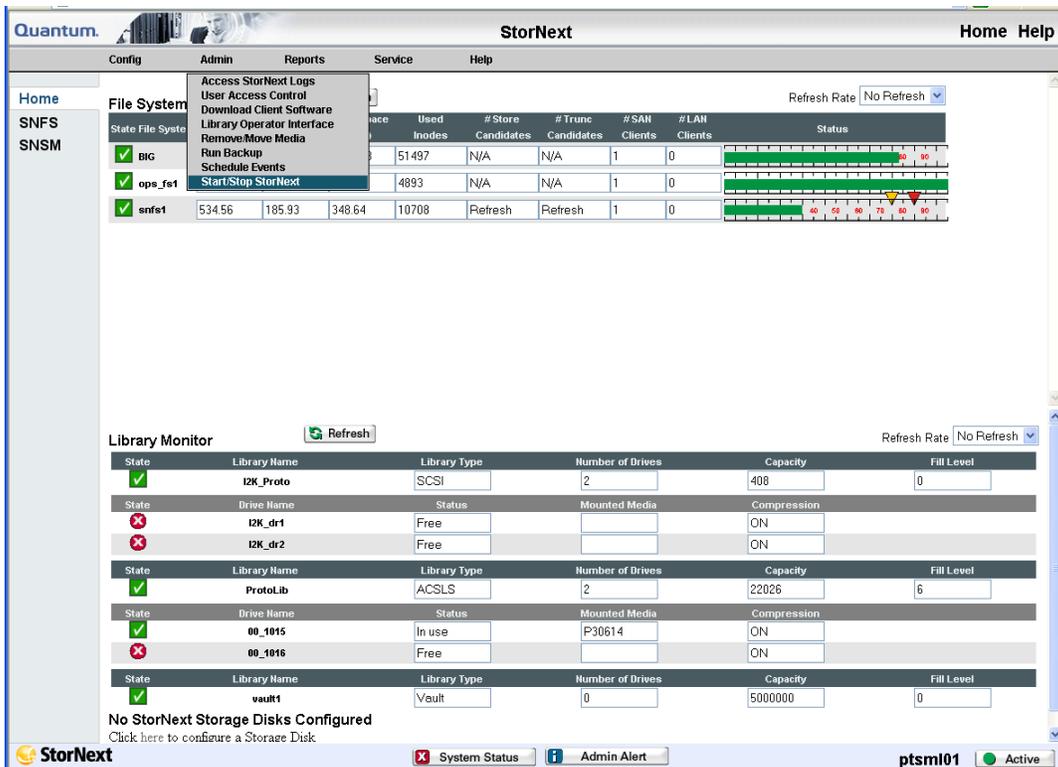


Figure 14.4-2. Admin Pull-Down Menu

- 5 Select **Start/Stop StorNext** from the **Admin** pull-down menu.
 - The **Start/Stop StorNext** page will be displayed (Figure 14.4-3).

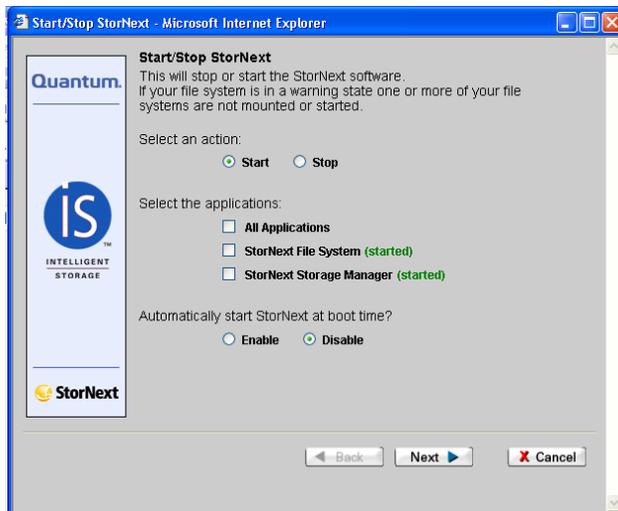


Figure 14.4-3. Start/Stop StorNext Page

- 6 Select the **Start** option under **Select an action**.
 - 7 Select the **Disable** option to disable the feature that automatically starts SNFS upon reboot.
NOTE: NEVER Select Automatically Start StorNext Manager at boot time.
 - 8 Select the **Next** button to save the changes and proceed.
 - 9 Click **Close** button when the status window displays **Success**.
-

14.4.2 Stopping the StorNext Application

To shutdown the StorNext System, both the server and its clients must be stopped. The clients must all be stopped first.

14.4.2.1 Stop the StorNext Clients

- 1 Log in as root into each **StorNext** client.
 - 2 To stop the **StorNext** clients, type:
`# service cvfs stop`
 - 3 Check to ensure client has been stopped:
`# ps -ef | grep cvfs`
NOTE: MAKE SURE THAT ALL CLIENTS ARE STOPPED.
 - 4 To Stop the StorNext Metadata server, log in as root (system administrator) into the active **StorNext Metadata** server (x4smvaa).
 - 5 To stop **StorNext** server, type:
`# service cvfs stop`
 - 6 Check to ensure server has been stopped
`# ps -ef | grep cvfs`
-

14.4.2.2 Disabling the Tape Archive System

- 1 From the **StorNext Home Page**, select **Stop/Start StorNext** from **Admin** pull down menu.
 - The **Start/Stop StorNext** page will be displayed (Figure 14.4-4).

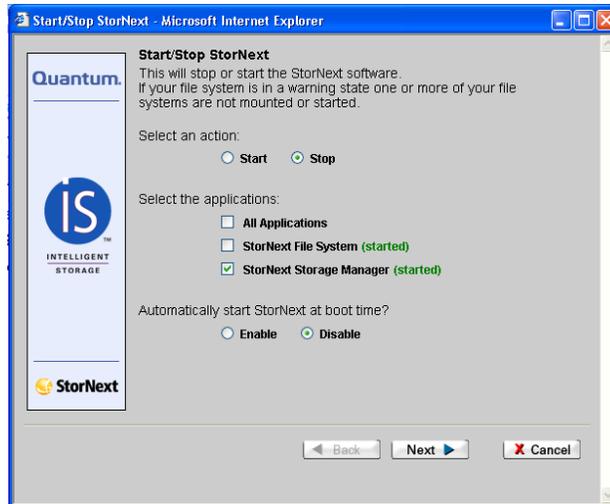


Figure 14.4-4. Stop StorNext Page

- 2 Select the **Stop** option from the **Select an Action** section.
- 3 Select **StorNext Storage Manger** checkbox from the **Select the components** section.
- 4 Select the **Disable** option.

NOTE: NEVER select **Automatically start StorNext at boot time?** EED has provided a script in the init.d directory to perform this action.

- 5 Select the **Next** button.
- 6 Click **Finish** when the screen displays **Success**.

14.4.3 Rebooting the StorNext Metadata Servers

The StorNext Metadata Servers (x4sml01, x4sml02) may need to be rebooted during certain anomalous conditions (e.g., system "hang," interruption of communication between StorNext and ACSLS, a required daemon is down).

In order to reboot StorNext Metadata Servers the operator must have root privileges. The following procedure demonstrates the steps to reboot StorNext Metadata Servers:

14.4.3.1 Rebooting the StorNext Metadata Servers

- 1 To reboot the **StorNext System**, you must stop both the Server and its Clients refer to Section 14.4.2 – **Stopping the StorNext Application**.
- 2 Perform Required Maintenance on StorNext Metadata Server.

3 Re-Start the StorNext Server and Clients (refer to Section 14.4.1 – Starting the StorNext Application).

14.4.4 Avoiding Loss of LUN Labels When Installing Red Hat

When installing Red Hat Enterprise Linux (RHEL) 5, the “Anaconda” installation program assumes that it owns any local or Storage Area Network (SAN) LUN (Logical Unit Number) that it can find and will **re-label** each LUN using Linux headers. If an installation is attempted while connected to the SAN, all of the StorNext LUN headers will be rewritten with Linux headers. StorNext or any other SAN filesystem that the fabric connection allows will cease functioning. Also, it is possible for Anaconda to target the incorrect disk and overwrite data on a SAN LUN. This will occur during a manual install via CD/DVD or an automated (kickstart) install. There is an undocumented option, **-ignore disks**, but it does not work when combined with any other kickstart options.

Warning: It is crucial that before installing Red Hat Linux (via CD, DVD, or kickstart) that any non-OS disks are removed from visibility of the server. For locally installed disks, disconnect or remove the drives. If the OS disk is on the SAN (diskless system), then any non-OS disks visible to the target host must be removed by one of the following methods:

- disconnecting the fiber cable,
- disabling the port on the fabric,
- using zoning,
- LUN masking,
- Navisphere or other storage management method.

When scanning devices on the target systems HBA, the only device that should be actively visible is the target OS disk. If installing Red Hat Linux on a local hard drive while attached to a FC SAN, simply disconnect the FC HBA connections to the SAN to easily protect the SAN. Additionally, a current list of StorNext Labels should be created and updated as required.

Do not grow the root device using logical volumes with in Linux. At the hardware level, these will be seen as two devices. During kickstart, only one will be visible. If for any reason the root disk must grow, and the root device is on the SAN, then grow the LUN at the hardware level. If the root device is local, then migrate the data to a larger drive.

14.4.4.1 Create a Listing of StorNext Labels

NOTE: Prior to installing Red Hat Enterprise Linux (RHEL) 5, create a file that contains the StorNext Labels (see the following procedure). Additionally, it is crucial that the Linux system’s SAN fiber channel (FC) cables be physically disconnected before attempting an install.

- 1 Log on as root to a host (Linux or other) that has persistent binding of the SAN LUNs.
- 2 Make a listing of the StorNext labels currently configured by using this command example:

```
# /usr/cvfs/bin/cvlabel -c > /labels
```

 - This will create a file called **labels** that contains the label, the device, size, VTOC label and comments.
- 3 Edit the **labels** file and **REMOVE ENTRIES THAT ARE NOT STORNEXT!**
 - If non-StorNext entries are still in the file, they will be given StorNext headers and will no longer be able to do their normal function.

For example, the following line is the boot (local) disk and should be deleted:


```
/dev/sda [MegaRAIDLD 0 RAID5 69G1.92] MBR Sectors: 0. SectorSize: 512
```
- 4 Copy the file to /usr/cvfs/config by using the following command:

```
# cp /root/labels /usr/cvfs/config
```
- 5 Logoff from StorNext.

If the StorNext headers are lost, perform the following procedure:

- 1 As root, login to the host where the labels have been copied (in step 1 above).
 - 2 Change directory to the cvfs configuration directory by entering the following:

```
# cd /usr/cvfs/config
```
 - 3 Copy the labels file to cvlabels by entering the following:

```
# cp labels cvlabels
```
 - 4 Run the cvlabel command by entering the following:

```
# /usr/cvfs/bin/cvlabel
```
 - 5 Check that the headers are correct using the following command:

```
# /usr/cvfs/bin/cvlabel -l
```
 - 6 Remove the cvlabels file using the command:

```
# rm /usr/cvfs/config/cvlabels
```
-

14.5 Loading and Removing Archive Media from the Scalar library

The Scalar i6000 library is equipped with an import/export (I/E) station allowing cartridges to be inserted or removed from the library without interrupting operations. The I/E station is on the front of the control module. The I/E station has a capacity of 24 LTO cartridges located in four removable magazines.

The Scalar i500 library is equipped with an import/export (I/E) station. The I/E station is located on the front of the control module. In a 5U station, the I/E has a capacity of six cartridges within a removable magazine. The 9U has a capacity of 12 cartridges within two removable magazines.

Table 14.5-1 provides an Activity Checklist for Loading, and Removing Archive Media.

Table 14.5-1. Loading and Removing Archive Media -Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Manager	Loading Archive Media	(P) 14.5.1.1	
2	Archive Manager	Removing Archive Media	(P) 14.5.1.2	
3	Archive Manager	Recovering Files From 9940 Media Native StorNext Tapes	(P) 14.5.1.3	
4	Archive Manager	Recovering Files from LTO Tapes Media, Native StorNext Tapes	(P) 14.5.1.4	

14.5.1 Loading Archive Media

Loading of media is appropriate when there are relatively small numbers of media to be loaded. With automated loading, StorNext assigns each cartridge a unique volume number, then enters the volumes in its database and marks the volumes Online in the database.

14.5.1.1 Loading Archive Media

- 1 Log in as **root** at the active **StorNext Metadata** server (**x4smvaa**). The **x** in the workstation name will be a letter designating your site: **m** = SMC, **l**=LaRC, **e**=LP DAAC, **n**=NSIDC (e.g., **n4smvaa** indicates a server at NSIDC).
- 2 Update the media file to add the appropriate volume information.

vi /usr/adic/MSM/internal/config/media_file_”library”

Format :

```
# [s]          any character in the set s, where s is a sequence of
#              characters and/or a range of characters, for example, [c-c].
#
# r*          zero or more successive occurrences of the regular expression
#              r. The longest leftmost match is chosen.
# Examples:
#
# ESY...      All six character labels that begin with ESY.
#
# [^0-9]..A*  All labels that do not begin with a digit, followed
#              by any 2 characters, followed by zero or more
#              occurrences of the character A.
#
# "DG" EF"    DG followed by double quote followed by a space
#              followed by EF
#
# Following is an example of what an entry in this file may look like:
#AML_1 ESY2..
#
#S2_98 SE925[0-2]
```

3 Place the Media in the Library. Select Config-Add Media from the StorNext Home page. The Add Media – introduction screen will appear (Figure 14.5-1).

4 Select the appropriate library media, then select the Next button.

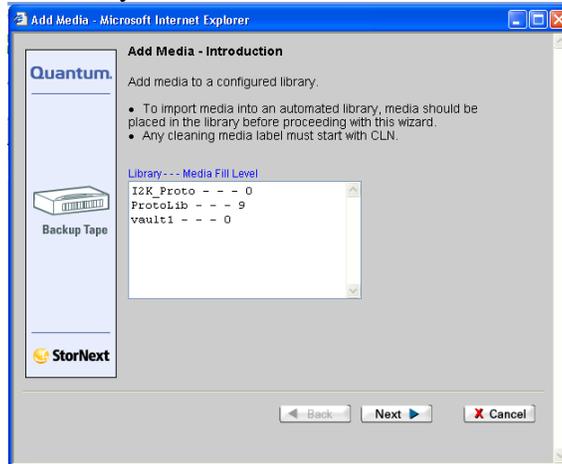


Figure 14.5-1. Add Media Page

5 Select the **Associated Library** (Figure 14.5-2), and press the **Next** button.

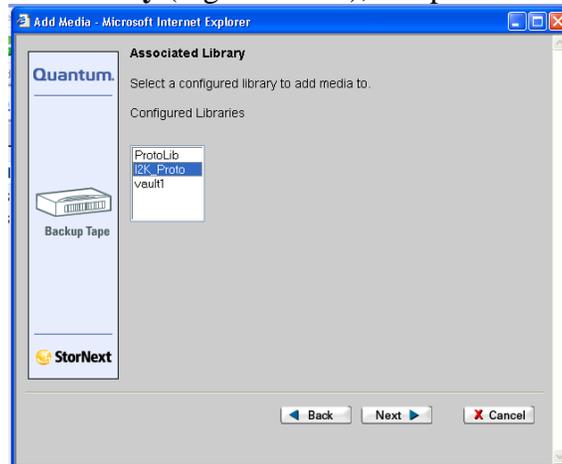


Figure 14.5-2. Associated Library Page

- 6 Select the **Bulk Load** button from the **Associated Library** page (Figure 14.5-3), and press the **Next** button.

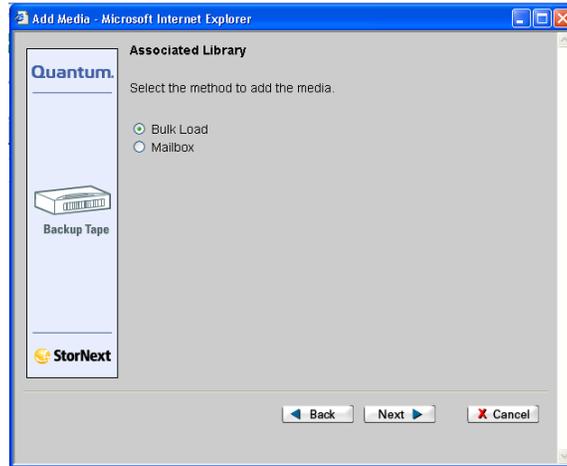


Figure 14.5-3. Associated Library Bulk Load Page

- The **Complete Add Media Task** screen displays.
- 7 Select **Next** from the **Complete Add Media Task** screen (Figure 14.5-4).
- The system will then automatically add your media.

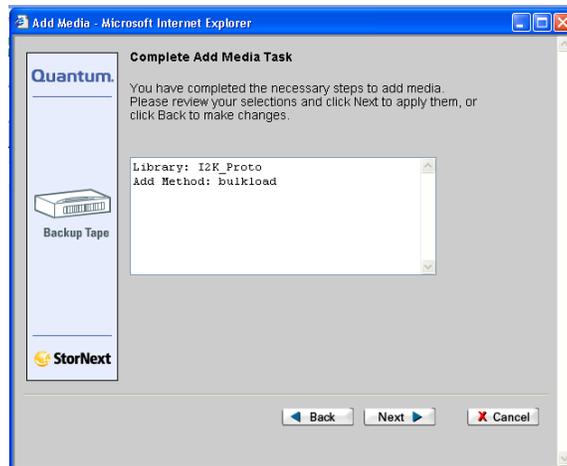


Figure 14.5-4. Complete Add Media Task Page

14.5.1.2 Removing Archive Media

- 1 From the **StorNext Home Page**, choose **Remove/Move Media** from the **Admin** pull down menu. (Figure 14.5-5)

The screenshot shows the Quantum StorNext Admin interface. The top navigation bar includes 'Config', 'Admin', 'Reports', 'Service', and 'Help'. The 'Admin' menu is open, showing options: 'Access StorNext Logs', 'User Access Control', 'Download Client Software', 'Library Operator Interface', and 'Remove/Move Media'. The 'Remove/Move Media' option is highlighted. Below the menu, there is a table with columns: 'State', 'File System', 'Used Inodes', '# Store Candidates', '# Trunc Candidates', '# SAN Clients', '# LAN Clients', and 'Status'. The table lists three file systems: 'BIG', 'ops_fe1', and 'snfs1'. Below this is the 'Library Monitor' section with a 'Refresh' button and a 'Refresh Rate' dropdown set to 'No Refresh'. The 'Library Monitor' table has columns: 'State', 'Library Name', 'Library Type', 'Number of Drives', 'Capacity', and 'Fill Level'. It lists three libraries: 'I2K_Proto' (SCSI), 'I2K_dr1' (Free), and 'I2K_dr2' (Free). Below this is another table with columns: 'State', 'Drive Name', 'Status', 'Mounted Media', and 'Compression'. It lists two drives: '00_1015' (Delayed) and '00_1016' (Free). At the bottom, there is a 'Vault' section with columns: 'State', 'Library Name', 'Library Type', 'Number of Drives', 'Capacity', and 'Fill Level'. It lists one vault: 'vault1' (Vault). The bottom status bar shows 'No StorNext Storage Disks Configured' with a link to configure a storage disk, and system status indicators for 'System Status' (red X), 'Admin Alert' (blue i), 'ptsmi01', and 'Active' (green dot).

Figure 14.5-5. Remove/Move Media Pull Down Menu

- 2 The **Remove Media or Move Media** screen will appear (Figure 14.5-6). Select the **Remove Media** button.
- 3 Select the appropriate **Library** and **Media Type**, then select the **Next** button.

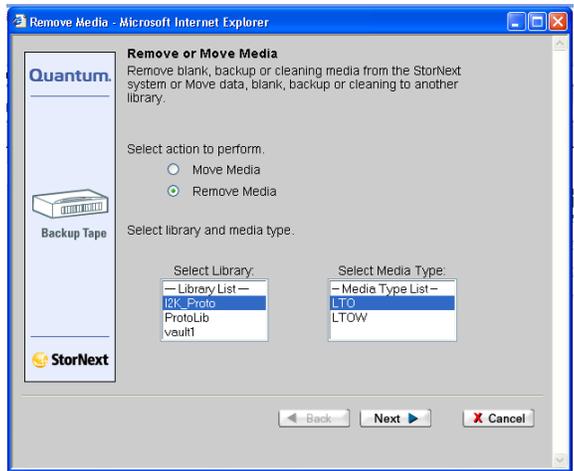


Figure 14.5-6. Remove or Move Media Page

4 Select the **Media** to be removed (Figure 14.5-7), then select the **Next** button.

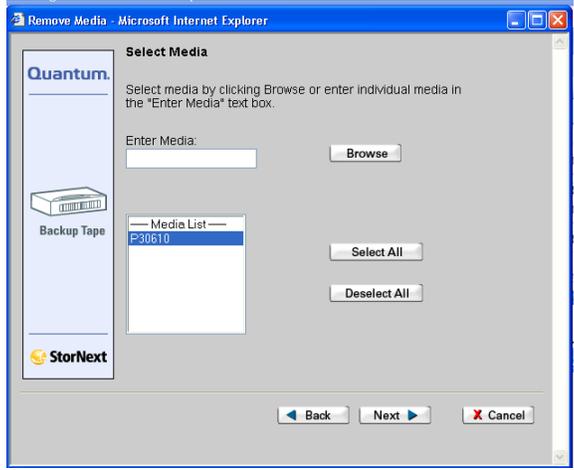


Figure 14.5-7. Select Media Screen

5 The **Completed Remove/Media Task** will appear (Figure 14.5-8).

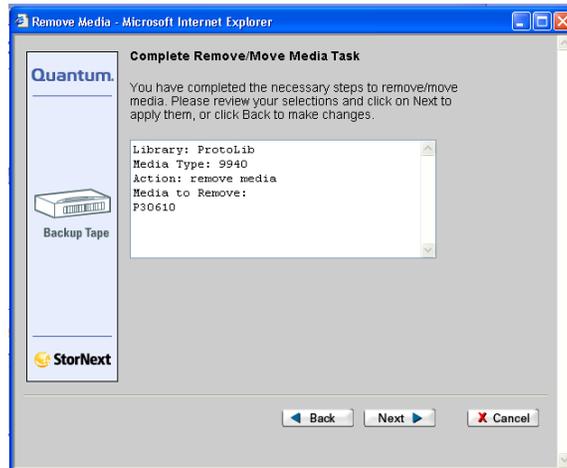


Figure 14.5-8. Complete/Remove Media Task Page

- 6 When the status screen indicates that the media has been removed, select **Finish**. The **Library Operator Interface (LOI)** page will appear.
- 7 Open the recessed latch on the **Cartridge Access Port (CAP)** door and remove the tape(s).
- 8 Update the media file to add the appropriate volume information. Type in:

```
# vi /usr/adic/MSM/internal/config/media_file_”library”
```

Format :

```
# [s]      any character in the set s, where s is a sequence of
#          characters and/or a range of characters, for example, [c-c].
# r*       zero or more successive occurrences of the regular expression
#          r. The longest leftmost match is chosen.
# Examples:
# ESY...   All six character labels that begin with ESY.
# [^0-9]..A* All labels that do not begin with a digit, followed
#          by any 2 characters, followed by zero or more
#          occurrences of the character A.
# "DG" EF  DG followed by double quote followed by a space
#          followed by EF
# Following is an example of what an entry in this file may look like:
#AML_1 ESY2.
#S2_98 SE925[0-2]
```

14.5.1.3 Recovering Files From LTO Media, Native StorNext Tapes

- 1 Log onto the **x4smvaa** machine as root.
 - 2 If the media to be removed is still available to StorNext, make the tape unavailable by entering the following command:

```
# fschmedstate <mediaID>-s unavail
```
 - 3 Follow the steps in Section 14.4.2.1 to shutdown the StorNext servers and clients
 - 4 Connect to the Scalar library, using the approved web browser.

```
http://192.168.xxx.xxx
```
 - 5 Take the partition offline:
Click **View-> Views**, the **Manage Views** dialog box appears.
Click the button to the right of the partition, toggles **online** to **offline**.
 - 6 Eject the target media from the tape library to the I/E Station:
From the Scalar GUI **View** menu, click the name of the partition:
Click **Operations -> Export**.
 - 7 The **Export Media** dialog box appears with the list of cartridges in the partition. Select the corresponding check box in the leftmost column for each cartridge that is to be exported (up to the number of I/E slots).
Click **OK**.
 - 8 Retrieve the tape from the I/E Station.
 - 7 Bring the partition online:
Click **View-> Views**, the **Manage Views** dialog box appears.
Click the button to the right of the partition, toggles **offline** to **online**.
Disconnect from the Scalar GUI.
 - 10 Follow the steps in 14.4.1.1 to start the StorNext servers and clients.
- The site maintenance coordinator will open a Quantum Service Request, and return the tape to quantum for analysis.
-

14.6 Backing Up the StorNext Application

StorNext provides the capability to perform both full and partial backups of metadata and database information. Full backups create backups of the full database dumps, snapshots of the file system metadata, and software configuration information. Partial backups create backups of the database journal files, metadata journal files, and software configuration information. Backups in SNSM version 2.8 are now written to a managed file system and stored off to media. When the data is stored to tape, the files are truncated to save room on disk. This is different than

earlier releases where backup data was saved to a local disk before being written to tape. Backups are run in several different ways:

- **Automatically**
 - Nightly as configured through the Scheduler
- **Manually:**
 - From the command line by running `snbackup`
 - From the GUI

The *snbackup* command-line utility is used to create the backups. The usage of `snbackup` can be found in the man page, and the utility also incorporates a `-h` option which can be used to display usage. This utility when first run will identify available system resources (a managed file system) to use for temporary storage of backup files. The largest capacity managed file system will be used for this temporary storage area. The selected file system will be stored in the */usr/adic/TSM/config/fs_sysparm* file.

This setting will be used in subsequent backups. Once all backup files have been created, the files are stored to media. The files are immediately truncated upon a successful store of all copies. This frees up the disk space that was borrowed from the managed file system. The number of copies, type of storage, and other attributes can be modified from the StorNext Home page and clicking **Admin > Backups** tab.

A complete set of backups is comprised of a full and all the subsequent partial backups. A set is identified by a unique identifier. This unique number is used during restores to tie all the backup elements together. Backups can be manually executed (command line or GUI) or they can be scheduled.

NOTE: By default, full backups are scheduled on Sundays. Partial backups are scheduled every day of the week except Sunday. When a backup completes, an e-mail notification is sent. The e-mail sent contains information about the backup. This backup information must be retained in order to successfully do a restore of the system in case of failure. If storage disks are used, the path of the storage disk media is shown.

IMPORTANT

DO NOT ATTEMPT RESTORE FROM THIS TAPE, OR ANY BACKUP UNLESS AUTHORIZED BY A CERTIFIED QUANTUM STORNEXT SUPPORT ENGINEER. EED SUSTAINING ENGINEERING DOES NOT SUPPORT ANY SYSTEM RESTORATION THAT HAS NOT BEEN APPROVED, OR PERFORMED SOLELY BY A QUANTUM CUSTOMER SUPPORT ENGINEER.

The following procedures describe how to run a manual backup. These backups are scheduled by default to run once a day. If a full backup already exists, you have the option to run either a full or partial backup.

By default, a full backup is run once a week. Weekly backups should include:

- The StorNext database.
- Configuration files.
- File system metadata dump file (after journal files are applied).

A partial backup runs on all other days of the week (that the full backup is not run). On other days backup should include:

- StorNext database journals.
- Configuration files.
- File system journal files.

Table 14.6-1 provides an Activity Checklist for StorNext Backup procedures addressed in this section.

Table 14.6-1. StorNext Backup Procedures - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Manager	Executing a StorNext Backup	(P) 14.6.1	
2	Archive Manager	Scheduling a StorNext Backup	(P) 14.6.2	

14.6.1 Executing a StorNext Backup

- 1 Connect to the StorNext web page using Firefox or Internet Explorer.
- 2 From the StorNext home page click Admin > Run Backup.
 - The options for the Admin drop-down menu (Figure 14.6-1) enable you to control day-to-day operations of StorNext. The Admin menu contains these options:

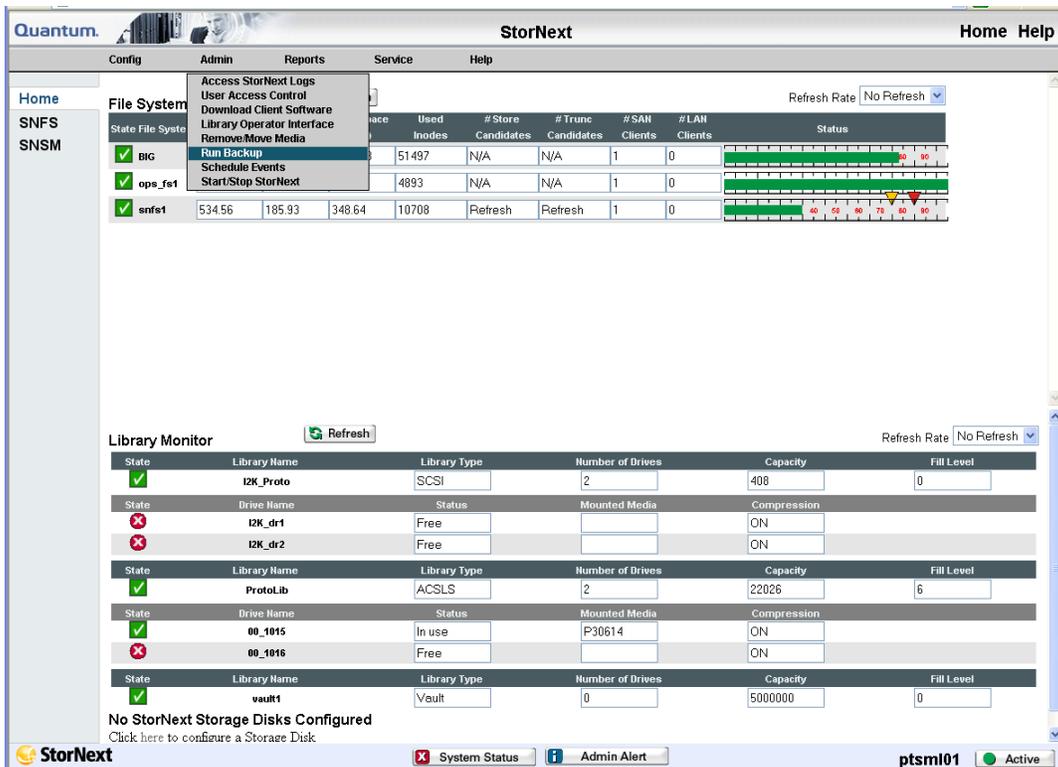


Figure 14.6-1. StorNext Admin Pull- Down Screen

- **Access StorNext Logs:** Access logs of StorNext operations
- **User Access Control:** Control user access to StorNext tasks
- **Download Client Software:** Download SNFS client software
- **Library Operator Interface:** Enter or eject media from the Library Operator Interface
- **Remove/Move Media:** Remove media from a library or move media from one library to another
- **Run Backup:** Run a backup of StorNext software
- **Schedule Events:** Schedule file system events including Clean Info, Clean Versions, Full Backup, Partial Backup, and Rebuild Policy
- **Start/Stop StorNext:** Start or stop the StorNext components

- 3 Select Run Backup. The Backup StorNext screen appears (Figure 14.6-2).

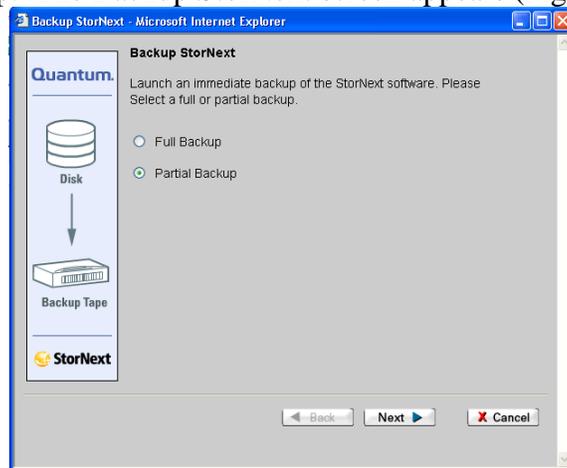


Figure 14.6-2. Backup StorNext Screen

- 4 Select the type of backup you want run, Full or Partial, then click Next. The **Complete Backup Task** screen appears (Figure 14.6-3).

NOTE: These backups DO NOT backup user data.

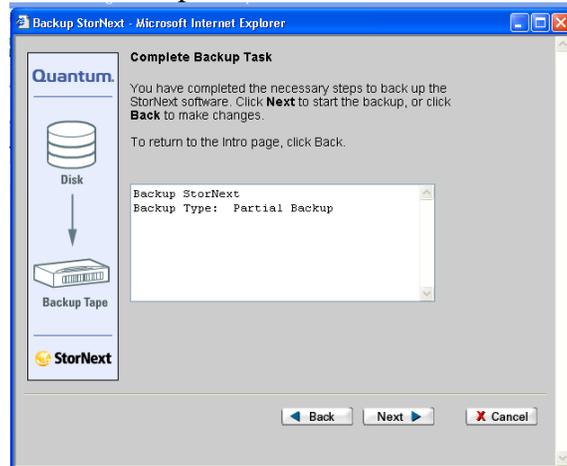


Figure 14.6-3. Complete Backup Screen

- 5 Click Next to start the backup.
 - 6 Click Finish when the Status screen displays success.
-

As stated previously, by default the StorNext Full Backup is set to execute once a week, and Partial Backups are performed on each day of the week that the full backups does not run on. To

schedule a backup outside of the default setting, use the Scheduling StorNext Events screen. You can use this screen to schedule all StorNext events. The following is an explanation of how to schedule a new event, such as backups.

Events that can be scheduled are:

- **Clean Info:** Scheduled background operation for removing knowledge of media from StorNext.
- **Clean Versions:** Clean old inactive versions of files.
- **Full Backup:** By default, a full backup is run once a week to back up the entire database, configuration files, and the file system metadata dump file.
- **Partial Backup:** By default, a partial backup is run on all other days of the week (that the full backup is not run). This backup includes database journals; configuration files, and file system journal files.
- **Rebuild Policy:** **Rebuild** the internal candidate lists (for storing, truncation, and relocation) by scanning the file system for files that need to be stored.

NOTE: The **Scheduler** does not dynamically update when dates and times are changed greatly from the current setting. You must reboot the system to pick up the change.

Each of these events have a default schedules set, these procedures allow you to reconfigure the schedules to suit your system needs.

14.6.2 Scheduling a StorNext Backup

- 1 From the **StorNext Home Page**, click **Admin > Schedule Events**.
 - The **Feature Schedules** screen appears (Figure 14.6-4).

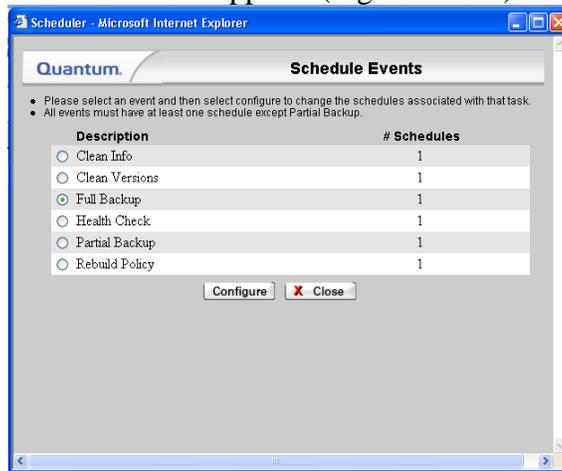


Figure 14.6-4. Feature Schedules Screen

- 2 Select a feature to schedule and click **Configure**.
 - The Feature Schedules screen displays the selected Feature and its current schedule (Figure 14.6-5).

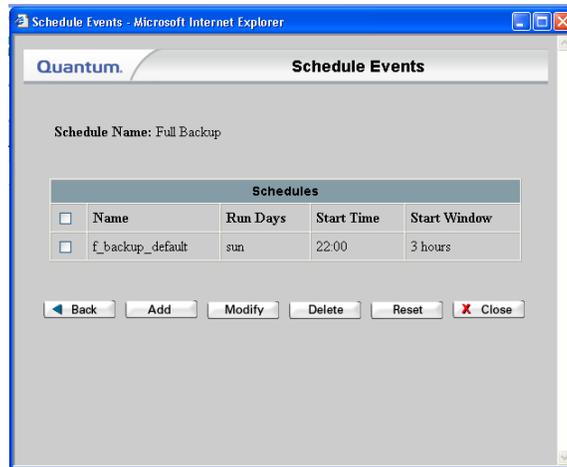


Figure 14.6-5. Selected Feature Schedules Screen

Select a **schedule**, then click one of the following:

- **Back:** Go back to the previous screen.
 - **Add:** Add a new schedule.
 - **Modify:** Change an existing schedule.
 - **Delete:** Delete an existing schedule.
 - **Reset:** Reset the schedule to the default settings.
 - **Close:** Close the window.
-

14.7 Scalar Library

14.7.1 Scalar I500 library

The operator panel is the touch screen display device located on the access door of the control module. The library operations and service functions are performed from this screen and from a remote web client. Both are required, since not all functionality is available through both. (Figures 14-7.1 and 14-7.2) The user interface has the following areas:

Header Bar – appears on every screen with the home, help and logout buttons.

Title Bar – is on the operator panel and gives library and partition panels.

Menu Bar – lists the menu choices on the web client only.

Main – Body of the screen.

Health/Navigation - Displays the status of the Library, Dives, and Media.

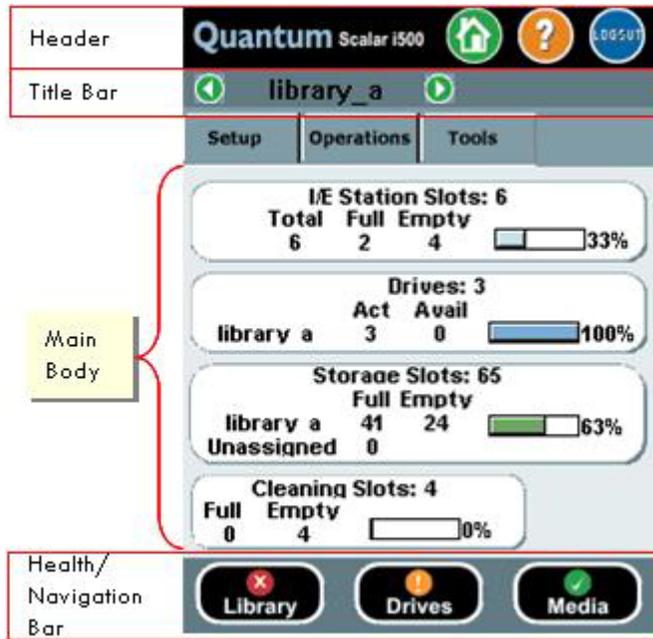


Figure 14.7-1. Scalar i500 Operator Panel User Interface

The buttons in the Health/Navigation Bar provide quick access to the information about the library. The buttons show RAS tickets that are reported by the library. Green state means that no tickets exist, yellow means there are open and unopened low and high tickets, red shows open and unopened urgent tickets.

The Capacity View is the default and shows the partitions, slots and drives in the main body of the screen.

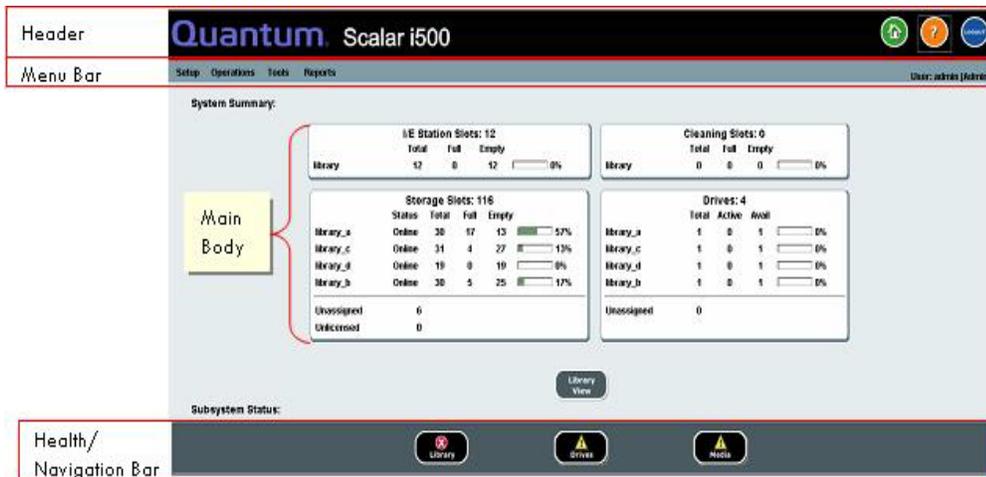


Figure 14.7-2. Scalar i500 Web Client User Interface

Table 14.7-1. StorNext Backup Procedures - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Manager	Scalar i500 Common Library Functions	(P) 14.7.1.1	
2	Archive Manager	Importing and Exporting Media	(P) 14.7.1.2	

14.7.1.1 Scalar i500 Common Library Functions

All users must login to the library to perform library functions or view the library operations, either through the operator panel or through the web browser.

- 1 Enter the **Username** and **Password** in the text boxes. (Figure 14.7-3) After initial setup the password is no longer the default, and must be obtained from the site administrator before continuing.

Note: To replace or reset your password, contact Quantum for technical support.



Figure 14.7-3. Scalar i500 Login Screen

- 2 Remember to always log out when library access is complete. With the web browser click the **Logout** button, or with the operator panel select **Operations > Logout**.
-

14.7.1.2 Importing and Exporting Media

WARNING: If the selected partition is online, it must be taken offline before the import or export operation is performed. This WILL impact operation.

The Import Export (IE) slots allow media to be imported and exported from the library. During import, the library's scanner automatically reads the barcode on new cartridges in the IE slots. Before importing cartridges, verify that all tape drives are unloaded and all cartridges are in the appropriate storage slot locations.

- 1 From the front of the library, insert cartridges into the IE station door. Once the door is closed the **Assign IE** screen appears on the operator panel. Select the partition and the slot to assign the cartridges. Select **Apply**.
- 2 Use the Import Media screens: **Operations>Media>Import** on either the operator panel or the web client to import the tapes into the partition. The partition will be taken off line and brought back online, so make sure StorNext is down on the server.

Note: The media must be configured into StorNext before it can be used. The instructions are found in Section 14.5.1.1 to of this document.

- 3 The Export media operation enables data cartridges to be exported from storage slots to the empty IE slots for removal from the library. From the menu select **Operations>Media>Export**. Provide the Partition and the tape cartridge that is to be removed. The partition will be taken offline and will be return online when export is complete.
-

14.7.2 Scalar I6000 library

The operator panel on the i6000 library is located on the front of the control module. It includes an indicator panel for the Robotics, Status and Power and a touch screen. The touch screen is the library navigation point and provides access to the Library Management Console (LMC). (Figure 14.7-4)

The LMC consists of five primary areas:

Title Bar – contains the library name

Menu Bar – provides menu access to all LMC commands

Tool Bar – quick access to most commonly executed functions

Library information panel – real-time library information

Overall System Status - gives real-time status information for the six subsystems of the physical library.

Table 14.7-2. StorNext Backup Procedures - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Manager	Scalar i2000 Common Library Functions	(P) 14.7.2.1	
2	Archive Manager	Importing and Exporting Media	(P) 14.7.2.2	

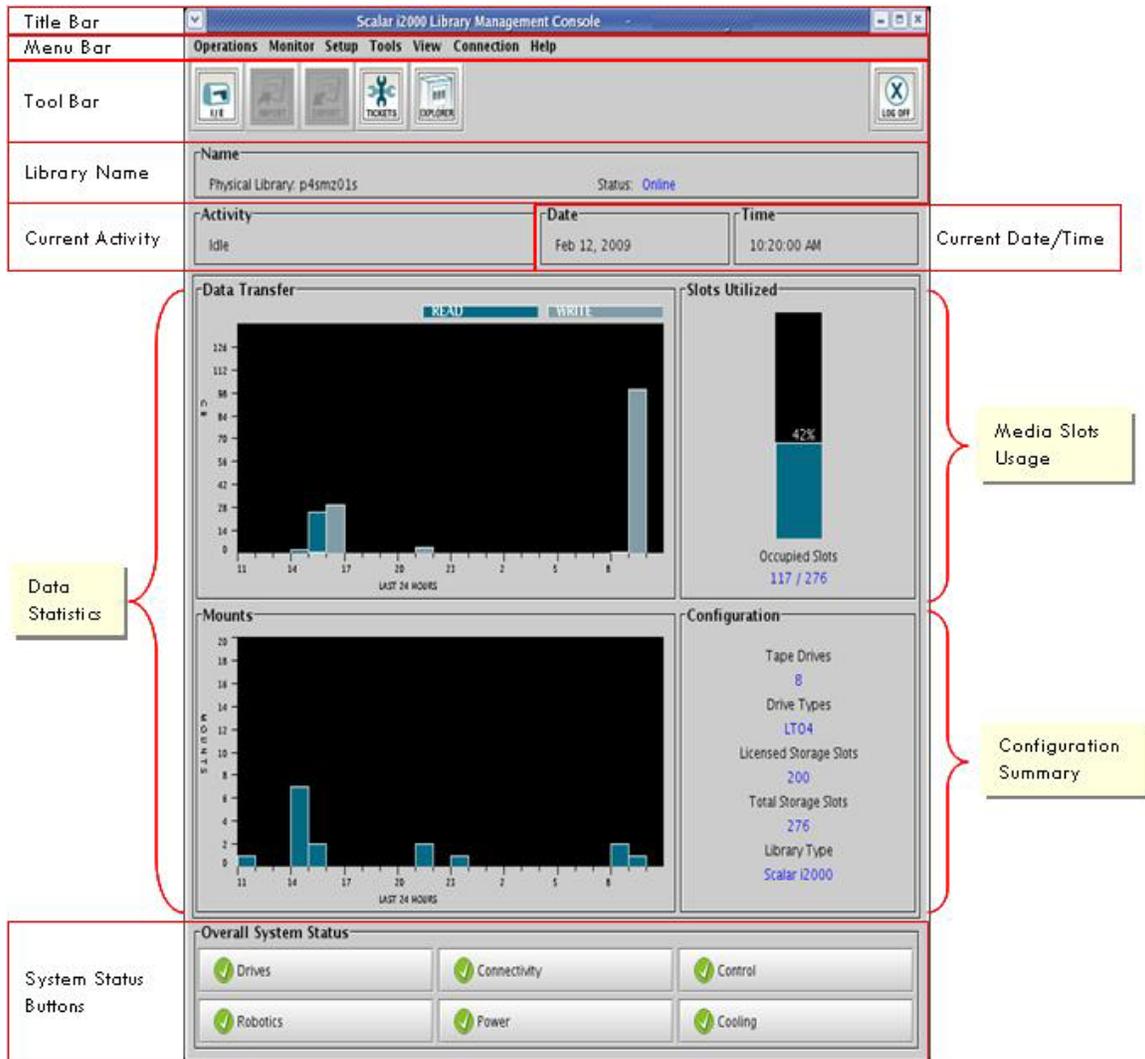


Figure 14.7-4. Scalar i6000 Library Management Console

14.7.2.1 Scalar i6000 Common Library Functions

- 1 From the **Tools** drop down menu, click **Library Explorer**. The Library Explorer dialog box appears.

- You can use the Library Explorer feature (Figure 14.7-5) to view a graphical presentation of all the drives, cartridges, and slots in the library. The Library Explorer can display all library elements according to physical location in any configuration.
- The Library Explorer features are available to administrator and service users, along with non-administrative users who have limited access to library functions. Users who do not have administrative privileges can perform all Operations options available to non-administrative users directly from the Library Explorer dialog boxes.



Figure 14.7-5. Scalar i6000 Library Explorer Screen

- 2 You can display library data using either the **Select Filter** options or clicking on a particular module in the **Select Module** area.
 - In the **Select Filter** area, search for and display specific criteria by device type and location coordinates, (or by Media ID.)
 - Select the **Device Type** filter, then from the **Type** drop-down list, click the appropriate device type: IE (I/E Station), Storage, or Drive. Click **Show**. The Module dialog box displays a graphical view of the library elements according to your Type filter choices.
 - To search for a specific cartridge according to the cartridge's barcode, select the **Media ID** filter, type the barcode in the **Media ID** field, then click **Show**. The Module dialog box displays the specific cartridge highlighted in red within the module where it is located.

- In the **Select Module** area, you can select a **specific module** in your library to view. On a multi-module library, all modules are represented.
 - In the **Select Module** area, click the **module** to view. The Module dialog box displays the current configuration of Rack one and Rack two according to the module you chose.
- 3 If you choose to search for an element by its address or choose to locate a cartridge by its media barcode, your search result appears in red in the Library Explorer Module dialog box.

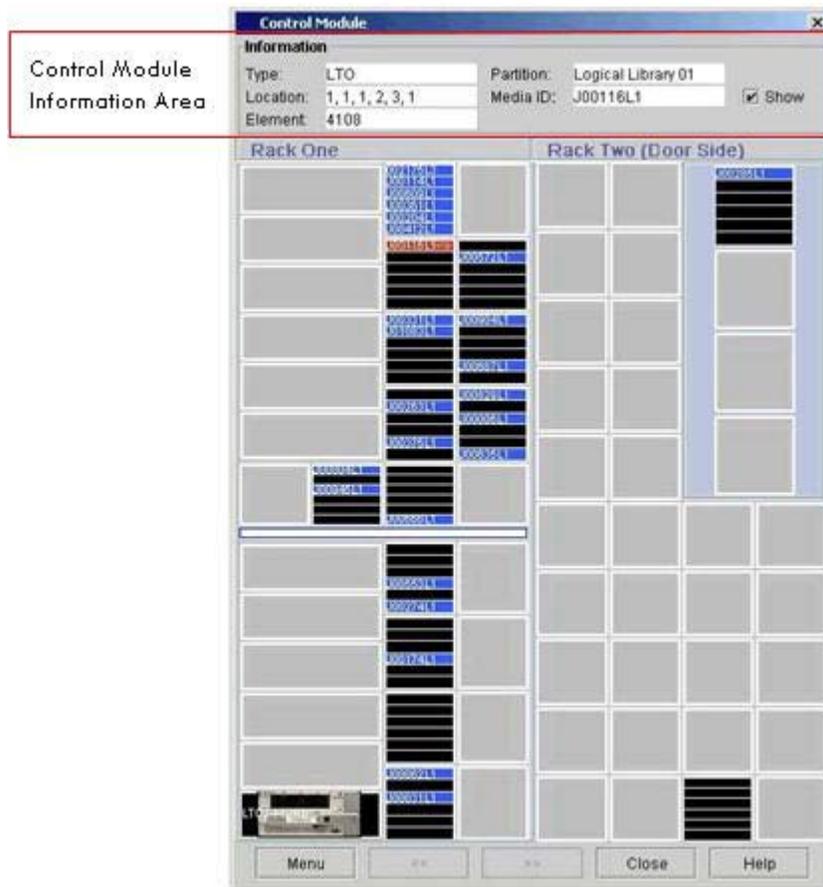


Figure 14.7-6. Scalar i6000 Control Module Information Screen

- 4 You can access Library Explorer Module from both the physical and partition views, but the functionality in the physical view is limited. If you are in a partition view, you can view slots and drives pertaining to that particular partition.
- The **Control Module** (Figure 14.7-6) dialog box displays the current configuration of the Rack.

- Slots containing cartridges are blue. Empty slots are black. Your search result appears in red.
 - Details concerning the particular cartridge, drive, or slot appear in the **Information area**.
 - Barcode numbers appear on slots containing cartridges.
 - If you click on a specific slot or drive, that slot or drive is highlighted in red, and details about the slot or drive appear in the Information area.
 - If you move the mouse over a specific segment in the module, a tool tip displays the coordinates of that particular segment.
 - To move from one module to another, click on the arrows at the bottom of the dialog box.
- 5** You can access Library Explorer Module from both the physical and partition views, but the functionality in the physical view is limited. If you are in a partition view, you can view slots and drives pertaining to that particular partition.
- The **Control Module** dialog box displays the current configuration of the Rack.
 - Slots containing cartridges are blue. Empty slots are black. Your search result appears in red.
 - Details concerning the particular cartridge, drive, or slot appear in the Information area.
-

14.7.2.2 Importing and Exporting Media

When you first start using your library, open the door and manually insert, directly into storage slots. The cartridges will not go back all the way if they are inserted incorrectly.

WARNING – StorNext must be shutdown to add media to the library, since this will take the partition offline.

The **Import Media** dialog box is used to add cartridges without interrupting library operations. Place cartridges in the I/E station. The scanner automatically reads the barcodes on new cartridges.

- 1** Make sure that you are viewing the partition into which you want to import a data cartridge. From the **View menu**, click the name of the appropriate partition.
- 2** Insert a **data cartridge** into an appropriate I/E station. You can insert multiple cartridges up to the maximum number of slots in your I/E station.
To see which I/E stations are associated with a particular partition, click **Monitor, IE Station**.
- 3** Click **Operations, Import** (or click the Import toolbar button).

WARNING – If the partition is not offline, you receive a message that asks you whether you want to take it offline. Click **Yes**.

The **Import Media** dialog box (Figure 14.7-7) appears with a list of cartridges in the I/E station displayed.

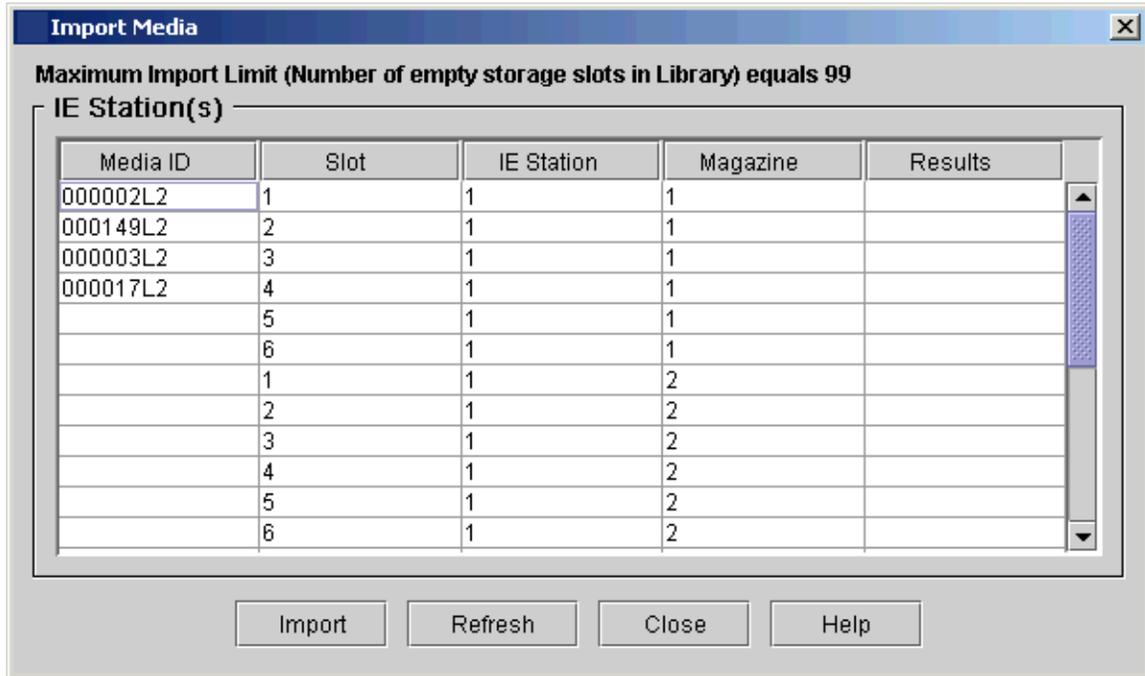


Figure 14.7-7. Scalar i6000 Import Media Screen

- 4 Select a cartridge (to highlight it), then click **Import**.

The picker automatically moves the cartridge from the I/E station to the first available empty slot in that partition. You cannot manually specify the slot.

Exporting Cartridges

- 1 Select the partition from which you want to export a data cartridge. From the **View menu**, click the **name of the appropriate partition**.
- 2 Click **Operations, Export** (or click the **Export** toolbar button.)
 - The **Export Media** dialog box (Figure 14.7-8) appears with a list of cartridges in the partition displayed.

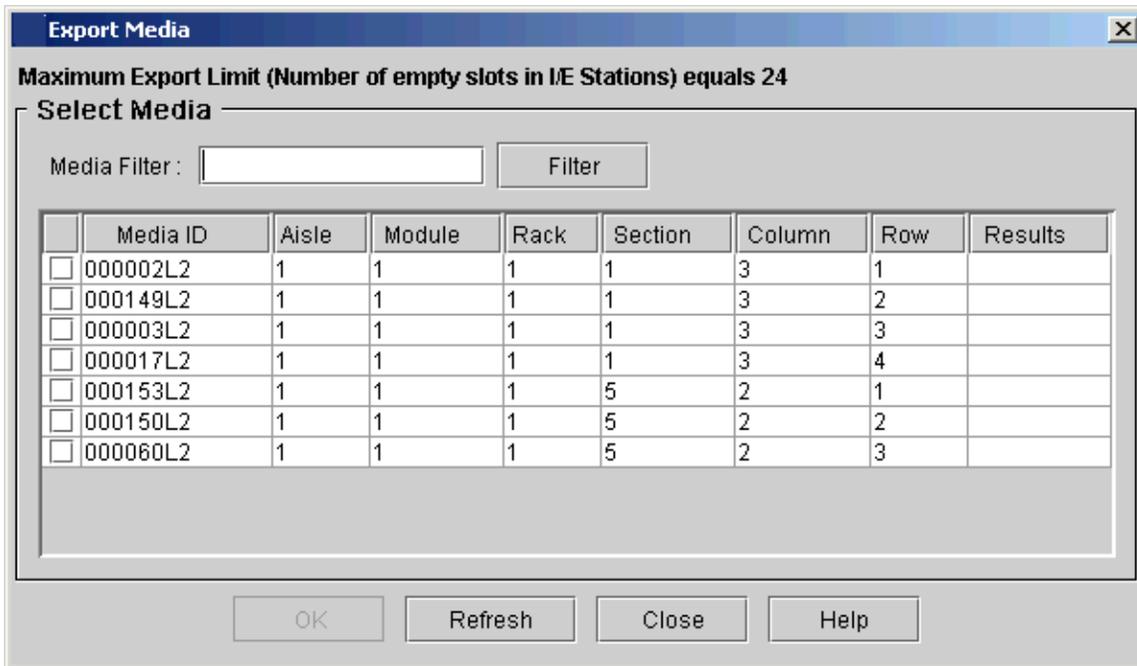


Figure 14.7-8. Scalar i2000 Export Media Screen

- 3 To display one or more media IDs that match a particular pattern, type a media filter in the **Media Filter** text box, then click **Filter**.
Filter performs a search for media IDs that match a particular pattern. In the example, the media filter has been set to capture media IDs beginning with the string “J00”.
- 4 Select the corresponding **check box** in the leftmost column for each cartridge that you want to export.
The maximum number of slots that are available in the I/E station partition appears at the top of the table.
- 5 Click **OK**.
 - All designated cartridges are exported to the I/E station slots that are associated with the partition. After the operation completes, the library automatically refreshes information in the table.

14.8 LTO Tape Drives

The next section provides the procedures for cleaning LTO tape drives. Table 14.8-1 provides the activity checklist for cleaning LTO tape drives.

Table 14.8-1. Table Cleaning Procedure - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Manager	Cleaning LTO Tape Drives	(P) 14.8.1	

14.8.1 Cleaning LTO tape drives

Note: To perform this procedure, you must have a least one cleaning tape in the library.

- 1 From the SNMS home page, choose **Drive, Clean Drive** from the **Admin** menu. The **Clean Drive** screen appears (Figure 14.8-1).



Figure 14.8-1. Clean Drive Screen

- 2 Select from the **Select Drive List** the drive to clean, then click **Apply**. The **Clean Drive** status screen appears.
 - 3 After the Status screen notifies you that the drive has been cleaned successfully, click **Close**.
-

14.9 Archive Maintenance Tasks - Deleting Granules

The Granule Deletion Utilities are a set of command line utilities:

- EcDsBulkSearch
- EcDsBulkDelete
- EcDsBulkUndelete
- EcDsDeletionCleanup

They will provide the EED Operations Staff with the ability to search granules for deletion, logically delete granules (marked granule for deletion), undelete logically deleted granules and physically clean up deleted granules, by using a set of command line interfaces.

The deletion process can involve deleting the specified granules along with associated granules, as long as no other granules reference the associated granules (e.g., browse, PH, QA). The deletion process can also involve deleting the specified granules even if they are inputs to other granules.

There are two phases to the granule deletion process:

- **Phase 1, Logical Deletion:** Logical deletion [marking or flagging granules as “deleted” or as DFA (Delete From Archive) only in the AIM database]. For the first phase, a command-line Bulk Delete utility (EcDsBulkDelete.pl) responds to operator-specified criteria for the deletion of granules by "logically" deleting from the inventory (AIM database) those granules that meet the criteria. The granules are marked as “deleted” and can no longer be accessed, but their inventory entries are not removed yet. The logical “deletion” may specify the flagging of granule files to be deleted from the archive (*Delete From Archive*, or DFA) only, leaving the inventory record intact, or it may specify *Physical Deletion*, which entails removal of the inventory record from the database as well as removal of the files from the archive. For each granule to be physically deleted an entry is made in the DsMdDeletedGranules table of the Inventory Database with a time stamp recording the logical deletion time. If applicable, the DFAFlag is set for the granule’s entry in the DsMdDeletedGranules table. Flagging DFA granules involves changing the value of the DeleteFromArchive entry in the DsMdGranules table from **N** to **Y**.
- **Phase 2, Physical Deletion:** Physical deletion involves the actual deletion of marked/flagged granules from the inventory database with removal of XML metadata files from the Small File Archive and the Science Granules from the Online and Tape File Archives. The second phase is actual deletion from the inventory of the granules marked for physical deletion (not DFA only). Physical deletion occurs when the operations staff runs the Deletion Cleanup utility (EcDsDeletionCleanup.pl). The Deletion Cleanup utility removes all Inventory entries for that granule from the AIM db, the XML file for the granule is removed from the XML archive, and all data files for the granule are removed from the Online and Tape File Archives. Please note, Granule Deletion Phase II will skip granules which are in public Data Pool or on order. So, before the operator executes Phase II script, he/she should run Data Pool Unpublish Utility with –aim option to unpublish granules which are marked for deletion in the Inventory database.

Periodically, as sufficient data removal from the archive makes it appropriate, Operations may elect to reclaim the tape space and recycle archive tapes. StorNext software commands (e.g., *volcomp*, *volclean*, *volformat*, and *volstat*) are used for that purpose.

Table 14.9-1 provides an Activity Checklist for Deleting Granules from the Archive.

Table 14.9-1. Deleting Granules - Activity Checklist

Order	Role	Task	Section
1	Archive Manager/Database Administrator	Generating a GeoID File	(P) 14.9.1.1
2	Archive Manager/Database Administrator	Mark Granules for Deletion (Logical)	(P) 14.9.2.1
3	Archive Manager/Database Administrator	Undelete Marked Granules for Deletion (Logical)	(P) 14.9.3.1
4	Archive Manager/Database Administrator	Running the Data Pool Unpublish Utility with –aim option	(P) 14.11.13.1
5	Archive Manager/Database Administrator	Deleting Granules, Phase 2: Running the Deletion Cleanup Utility	(P) 14.9.4.1
6	Archive Manager/Database Administrator	Deleting Granules, Phase 2: Rerun unfinished Physical Cleanup	(P) 14.9.4.2

14.9.1 Generating a GeoID File

A GeoID is the granule identification portion of a Universal Reference (UR); it contains the BaseType, SubType (ESDT ShortName and VersionID) and databaseID. For example, the GeoID SC:AST_L1BT.001:5850 indicates BaseType SC (science granule), ShortName AST_L1BT (ASTER Level 1B thermal infrared data) VersionID 001, and databaseID 5850. The GeoID is different for each granule in the system.

GeoID files are input files for scripts used in deleting (or “undeleting”) ECS granules from the inventory, archive, or Data Pool. A GeoID file consists of a list of GeoIDs for granules that are to be deleted (or “undeleted”). One way to generate a file of granule GeoIDs is to use the Bulk Search utility (EcDsBulkSearch.pl), which allows the operator to specify criteria for selecting granules on the basis of certain characteristics (e.g., ESDT, version, and date inserted in the archive). Subsequently, the GeoID file can be used as input to the Bulk Delete utility, the Bulk Undelete utility, or the Data Pool Cleanup Utility. Table 14.9-2 provides a description of the parameters used in executing the Bulk Search utility.

Table 14.9-2. Command Line Parameters of the EcDsBulkSearch.pl (1 of 2)

Parameter Name	Mandatory	Description
name	No	ESDT Short Name of the granules to delete.
version	No	ESDT Version ID of the granules to delete.
begindate	No	<mm/dd/yyyy> <hh:mm:ss> Search only for granules whose BeginningDateTime is greater than or equal to the specified date and time.

Table 14.9-2. Command Line Parameters of the EcDsBulkSearch.pl (2 of 2)

Parameter Name	Mandatory	Description
enddate	No	<mm/dd/yyyy> <hh:mm:ss> Search only for granules who's EndingDateTime is less than or equal to the specified date and time.
acquirebegin	No	<mm/dd/yyyy> <hh:mm:ss> Search only for granules whose BeginningDateTime is greater than or equal to the specified date and time. This option is the same as '- begindate', except that it can be combined with 'acquireend' and used in a 'BETWEEN' clause.
acquireend	No	<mm/dd/yyyy> <hh:mm:ss> Search only for granules who's BeginningDateTime is less than or equal to the specified date and time. This option is usually used in conjunction with 'acquirebegin'.
insertbegin	No	<mm/dd/yyyy> <hh:mm:ss> Search only for granules who's insertTime is greater than or equal to the specified date and time
insertend	No	<mm/dd/yyyy> <hh:mm:ss> Search only for granules who's insertTime is less than or equal to the specified data and time
localgranulefile	No	The name of a file containing Local Granule IDs to be converted into Geoids
geoidfile	Yes	Name of file containing geoids of the granules to delete.
physical	No	Search only for deleted granules.
dfa	No	Search only for DFA'd granules
mode	Yes	The ECS mode in which the program is to operate, this parameter can be omitted if the environment variable MODE is set.
user	Yes	The user ID for database login
limit	No	Search will return top of <n> granules specified by limit
server	Yes	The name of Sybase server. this parameter can be omitted if the environment variable SYB_SQL_SERVER is set
database	Yes	The name of the database, this parameter can be omitted if the environment variable SYB_DB_NAME is set
password	No	The name of the database login password, the utility will prompt user to enter the password if it is not specified in the command line (for security reason, not recommend to specify password in the command line)

Generic Bulk Search format:

EcDsBulkSearch.pl -name <shortname> -version <version ID> -begindate <mm/dd/yyyy> <hh:mm:ss> -enddate <mm/dd/yyyy> <hh:mm:ss> -insertbegin <mm/dd/yyyy> <hh:mm:ss> -insertend <mm/dd/yyyy> <hh:mm:ss> -acquirebegin <mm/dd/yyyy> <hh:mm:ss> -acquireend <mm/dd/yyyy> <hh:mm:ss> -DFA -physical

-localgranulefile <path/filename> **-geoidfile** <geoid file> **-limit** <granule limit>
-mode <ECS mode> **-server** <SYBASE_SERVER> **-database** <DB_NAME> **-user**
<db_user> **-password** <database login password>

14.9.1.1 Generating a Geoid File

- 1 Log in at the x4dpl01 host, where the Bulk Search utility is installed.
- 2 To change to the directory for starting the Bulk Search utility at the UNIX prompt enter:
cd /usr/ecs/<MODE>/CUSTOM/utilities
 - The working directory is changed to /usr/ecs/<MODE>/CUSTOM/utilities.
- 3 To set up relevant environment the following commands would allow the Bulk Search utility to run using the OPS mode Inventory database at the DAAC:
setenv MODE OPS
setenv SYB_SQL_SERVER x4dbl03_srvr
setenv SYB_DB_NAME EcInDb
 - The <ECS mode> value specified for the MODE parameter indicates the ECS mode (e.g., OPS, TS1, or TS2) to be searched.
 - If this environment variable is set, the **-mode** command line argument does not need to be given when starting the Bulk Search utility.
 - The <Sybase server> value specified for the SYB_SQL_SERVER parameter indicates the Sybase (database) server for the Inventory database.
 - If this environment variable is set, the **-server** command line argument does not need to be given when starting the Bulk Search utility.
 - The <database name> value specified for the SYB_DB_NAME parameter indicates which database (e.g., EcInDb, EcInDb_TS1, or EcInDb_TS2) is involved in the search.
 - If this environment variable is set, the **-database** command line argument does not need to be given when starting the Bulk Search utility.
- 4 Example 1:
 - Generate a file of GeoIDs deletion by shortname, versionid and inclusive temporal range:
EcDsBulksearch.pl -geoidfile </path/geofilename> -name <ESDT ShortName> -version <ESDT versionId> -begindate <mm/dd/yyyy> -user <db_userid>
- 5 Example 2:
 - Generate a file of GeoIDs for all MYD09GQ.001 granules marked “DFA” in the OPS mode.
EcDsBulkSearch.pl -DFA -name MYD09GQ -version 001 -password password -geoidfile MYD09GQ_Dec23.geoid
- 6 Example 3:
 - Generate a file of GeoIDs for all deleted (“physical” deletion) MYD09GQ_100 granules in the OPS mode at the DAAC.
EcDsBulkSearch.pl -physical -name MYD09GQ -version 100 -password <password> -geoidfile MYD09GQ_Dec23.geoid

- 7 When the Bulk Search utility has completed its run and the GeoID output file is available, at the UNIX prompt enter:
- vi** *<geoid file>*
- *<geoid file>* refers to the GeoID file to be reviewed
 - Although this procedure has been written for the **vi** editor, any UNIX editor can be used to edit the file.
- 8 Review the file entries to identify any problems that have occurred.
- The GeoID file must contain GeoIDs in the format *<BaseType>:<ESDT_ShortName.VersionID>:<databaseID>*.
 - For example:
SC:PM1ATTNR.077:2013496393
 - The GeoID in the example indicates BaseType SC (science granule), ShortName PM1ATTNR (AQUA attitude data in native format) VersionID 077, and databaseID 2013496393.
 - There may be no spaces or blank lines in the file.
- 9 Use UNIX editor commands to fix problems detected in the file.
- The following **vi** editor commands are useful:
 - **h** (move cursor left).
 - **j** (move cursor down).
 - **k** (move cursor up).
 - **l** (move cursor right).
 - **a** (append text).
 - **i** (insert text).
 - **r** (replace single character).
 - **x** (delete a character).
 - **dw** (delete a word).
 - **dd** (delete a line).
 - **ndd** (delete *n* lines).
 - **u** (undo previous change).
 - **Esc** (switch to command mode).
- 10 Press the **Esc** key.
- 11 At the **vi** prompt enter:
- ZZ**
- **vi** exits and the edited file is saved.
 - To exit **vi** without saving the new entries in the file type **:q!** then press **Return/Enter**.
 - UNIX command line prompt is displayed.
-

14.9.2 Deleting Granules, Phase 1: Mark Granules for Deletion (Logical)

Once granules have been identified/selected for deletion, the operator runs the Bulk Delete utility, a Perl script, **EcDsBulkDelete.pl**. There are two types of runs that can be performed with the Bulk Delete utility:

- 1 Physical.
- 2 DFA.

A “physical” deletion run results in marking granules in the geoidfile as logically deleted. Specifying “DFA” (not able to combine with physical) will mark the granules in the geoidfile as DFAed, meaning the metadata will be kept but the granule files will be removed from the archive.

As previously mentioned the Bulk Delete utility responds to operator-specified criteria for the deletion of granules by "logically" deleting from the Inventory Database those granules that meet the criteria. The granules are marked as “deleted” and can no longer be accessed, but their inventory entries are not removed yet. The logical “deletion” may specify the flagging of granule files to be deleted from the archive (*Delete From Archive*, or DFA) only, leaving the Inventory Database record intact, or it may specify *Physical Deletion*, which entails removal of the Inventory Database record from the database as well as removal of the files from the archive.

For each granule to be physically deleted an entry is made in the DsMdDeletedGranules table of the Inventory Database with a time stamp recording the logical deletion time. If applicable, the DFAFlag is set for the granule’s entry in the DsMdDeletedGranules table. Flagging DFA granules involves changing the value of the DeleteFromArchive entry in the DsMdGranules table from **N** to **Y**. Table 14.9-3 provides a description of the parameters used in executing the Bulk Delete utility.

Table 14.9-3. Command Line Parameters for EcDsBulkDelete.pl (1 of 2)

Parameter Name	Mandatory	Description
geoidfile	Yes	Name of file containing geoids of the granules for deletion.
physical	Yes if not dfa	Specifying this parameter will mark granules in the geoidfile logically deleted.
DFA	Yes if not physical	Specifying this parameter (not able to combine with physical) will mark the granules in the geoidfile as DFAed, meaning the metadata will be kept but the granule files will be removed from the archive.
delref	No	Optional. When given, indicates that non SC/LM granule should be deleted even if it is associated to undeleted SC/LM granules Note: This option has no effect on deleting SC/LM granules. They are always deleted regardless of being referenced or not.
noassoc	No	Optional. When given, indicates that associated granules (Browse granules etc.) will not be deleted.

Table 14.9-3. Command Line Parameters for EcDsBulkDelete.pl (2 of 2)

Parameter Name	Mandatory	Description
mode	Yes	The ECS mode in which the program is to operate, this parameter can be omitted if the environment variable MODE is set.
user	Yes	The user ID for database login
server	Yes	The name of Sybase server. this parameter can be omitted if the environment variable SYB_SQL_SERVER is set
database	Yes	The name of the database, this parameter can be omitted if the environment variable SYB_DB_NAME is set
password	No	The name of the database login password, the utility will prompt user to enter the password if it is not specified in the command line (for security reason, not recommend to specify password in the command line)
log	No	The name of the log file to which a deletion report will be written. If this is not provided, it will default to /usr/ecs/<MODE>/CUSTOM/logs/

NOTE: A prerequisite to deleting granules is having a file of GeoIDs (corresponding to granules) for use as input to the Bulk Delete utility. Although it is possible to manually create a file of GeoIDs, an easier way is to use the Bulk Search utility to generate a list of GeoIDs based on criteria specified when running the Bulk Search utility (refer to the procedure for running the Bulk Search utility.)

Generic Bulk Delete format:

EcDsBulkDelete.pl -physical | -DFA -delref -noassoc -user <db_user> -password <passwd>-geoidfile <path/filename>-mode <MODE> -server <SYBASE_SERVER> -database <DB_NAME>-log <log_file_name>

14.9.2.1 Mark Granules for Deletion (Logical)

- 1 Log in at the x4dpl01 host, where the Bulk Delete utility is installed.
- 2 To change to the directory for starting the Bulk Search utility at the UNIX prompt enter:
cd /usr/ecs/<MODE>/CUSTOM/utilities
 - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.
- 3 To set up relevant environment the following commands would allow the Bulk Search utility to run using the OPS mode Inventory database at the DAAC:
setenv MODE OPS
setenv SYB_SQL_SERVER x4dbl03_srvr
setenv SYB_DB_NAME EcInDb

NOTE: There are two types of runs that can be performed with the Bulk Delete utility; i.e., “physical,” or “DFA.” A “physical” deletion run results in marking each specified granule and metadata as deleted from both inventory and archive. A “DFA” run involves deletion from the archive only.

- 4 To mark granules for logical “physical” deletion (i.e., granules plus associated granules such as Browse, QA and PH), at the UNIX prompt enter:
- ```
EcDsBulkDelete.pl -physical -delref -noassoc -geoidfile <path/geoid_name> -user -log <log_file_name.log>
```
- The `-delref` option (i.e., delete granules that are referenced by other granules) indicates that a non-science/limited (non-SC/LM) granule should be deleted even if it is associated with “undeleted” SC/LM granules.
  - The `-delref` option has no effect on deleting SC/LM granules. They are always deleted regardless of whether or not they are referenced.
- 5 To mark granules for logical “DFA” deletion (i.e., meaning the metadata will be kept but the granule files will be removed from the archive), at the UNIX prompt enter:
- ```
EcDsBulkDelete.pl -DFA -geoidfile <path/geoid_name> -user -log <log_file_name.log>
```
- The `-DFA` option indicates that the granules listed in the GeoID file are to be marked as “Delete From Archive” only (does not involve a “physical” deletion).
 - The Bulk Delete utility records information about utility events in the log file.
- 6 When the Bulk Delete utility has completed its run and the log file is available, at the UNIX prompt enter:
- ```
more <log filename>
```
- The contents of the log file is displayed.
- 

### 14.9.3 “Undeleting” Granules from the Archive and Inventory

Logically deleted or DFAed granule(s) can be undeleted by `EcDsBulkUndelete` utility.

The **BulkUndelete** utility requires a **geoid** file, in which all granules intended to be undeleted are properly listed (i.e., “physical” or `-DFA`).

The following command line format is used to **undelete** granules which have been marked for “physical” deletion or DFA deletion.

```
EcDsBulkUndelete.pl -physical | -DFA -noassoc -user <db_user>-password <passwd>-geoidfile <path/filename>-mode <MODE> -server <SYBASE_SERVER> -database <DB_NAME> -log <log_file_name>
```

Table 14.9-4 provides a description of the parameters used in the Bulk Undelete utility.

**Table 14.9-4. Command Line Parameters for EcDsBulkUndelete.pl**

| Parameter Name | Mandatory | Description                                                                                                                                                                                                         |
|----------------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Geoidfile      | Yes       | Name of file containing geoids of the granules for deletion.                                                                                                                                                        |
| Physical       | Yes       | Specify this parameter will undelete granules specified in the geoid file which have been previously logically deleted.                                                                                             |
| DFA            | Yes       | Specify this parameter (not able to combine with physical) will un-DFA granules specified in the geoid file which have been previously marked as DFA.                                                               |
| Noassoc        | No        | Optional. When given, indicates that associated granules (Browse granules etc.,) will not be undeleted.                                                                                                             |
| Mode           | Yes       | The ECS mode in which the program is to operate, this parameter can be omitted if the environment variable MODE is set.                                                                                             |
| User           | Yes       | The user ID for database login                                                                                                                                                                                      |
| Server         | Yes       | The name of Sybase server. this parameter can be omitted if the environment variable SYB_SQL_SERVER is set                                                                                                          |
| Database       | Yes       | The name of the database, this parameter can be omitted if the environment variable SYB_DB_NAME is set                                                                                                              |
| Password       | No        | The name of the database login password, the utility will prompt user to enter the password if it is not specified in the command line (for security reason, not recommend to specify password in the command line) |
| Log            | No        | The name of the log file to which an undeletion report will be written. If this is not provided, it will default to /usr/ecs/<MODE>/CUSTOM/logs/                                                                    |

**NOTE:** A prerequisite to “undeleting” is having a file of GeoIDs (corresponding to granules) for use as input to the Bulk Undelete utility. Although it is possible to manually create a file of GeoIDs, an easier way is to use the Bulk Search utility to generate a list of “deleted” GeoIDs based on criteria specified when running the Bulk Search utility (refer to the procedure for running the Bulk Search utility).

### 14.9.3.1 Undelete Marked Granules for Deletion (Logical)

- 1 Log in at the x4dpl01 host, where the Bulk Undelete utility is installed.
- 2 To change to the directory for starting the Bulk Search utility at the UNIX prompt enter:  
**cd /usr/ecs/<MODE>/CUSTOM/utilities**
  - The working directory is changed to /usr/ecs/<MODE>/CUSTOM/utilities.
- 3 To set up relevant environment the following commands would allow the Bulk Search utility to run using the OPS mode Inventory database at the DAAC:  
**setenv MODE OPS**  
**setenv SYB\_SQL\_SERVER x4db103\_srvr**  
**setenv SYB\_DB\_NAME EcInDb**

**NOTE:** There are two types of runs that can be performed with the Bulk Undelete utility; i.e., “physical,” or “DFA.” A “physical undeletion” run results in removing

“deleted” markings for the granules/metadata in both inventory and archive. A “DFA undeletion” run involves removing “deleted” markings for the granules in the archive only.

4 To perform a “Physical” undeletion, at the UNIX prompt enter the following:

```
EcDsBulkUndelete.pl -physical -noassoc -geoidfile <path/geoid_name> -user
<database user ID >-log <log_file_name.log>
```

- The Bulk Undelete utility runs and removes “deleted” markings for the granules specified in the GeoID file in the archive.
- The Bulk Undelete utility records information about utility events in the log file.

5 To perform a “DFA undeletion,” at the UNIX prompt enter the following:

```
EcDsBulkUndelete.pl -DFA -geoidfile <geoid file>-user <database user ID> -log <log
filename>
```

- The -DFA option indicates that “deleted” markings are to be removed for the granules in the archive only.
  - The Bulk Undelete utility records information about utility events in the log file.
- 

#### 14.9.4 Deleting Granules, Phase 2: Running the Deletion Cleanup Utility

Once granules have been marked/flagged for deletion, the operator runs the Deletion Cleanup utility, **EcDsDeletionCleanup.pl**. As previously mentioned the Deletion Cleanup utility removes all inventory rows (in the Inventory Database) for “physically deleted” granules that were marked as “deleted,” including rows referencing related information (e.g., BR, PH, and QA). The database records for the granule(s) are removed from all appropriate Inventory database tables.

When the utility is executed, the utility checks for any unfinished work from a previous run(s). If found, the user will be prompted the following options:

- Rerun unfinished run only
- Start a new run which includes unfinished run(s)
- Quit

The EcDsDeletionCleanup requires user’s interactions during execution.

```
EcDsDeletionCleanup.pl -user <db_user>-mode <MODE> -server <SYBASE_SERVER> -
database <DB_NAME> -batch <number> -grbatch <number> -phbatch <number> -log
<log_file_name>
```

There are various command line parameters that are used in combination with each other. Table 14.9-5 provides a description of the parameters.

The operations staff can control the lag time between logical deletion and physical deletion. The lag time is entered into the Deletion Cleanup utility, which deletes inventory entries only for granules that have been logically deleted prior to that time period.

**Table 14.9-5. Command Line Parameters for EcDsDeletionCleanup**

| Parameter Name | Mandatory | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| mode           | Yes       | The ECS mode in which the program is to operate, this parameter can be omitted if the environment variable MODE is set.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| user           | Yes       | The user ID for database login                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| server         | Yes       | The name of Sybase server. this parameter can be omitted if the environment variable SYB_SQL_SERVER is set                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| database       | Yes       | The name of the database, this parameter can be omitted if the environment variable SYB_DB_NAME is set                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Log            | No        | The name of the log file to which an undeletion report will be written. If this is not provided, it will default to /usr/ecs/<MODE>/CUSTOM/logs/                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| batch          | No        | The batch size for populating DsStPendingDelete table in batch. This parameter can be omitted if the environment variable BATCH_SIZE_GRANULE is set. If the environment variable BATCH_SIZE_GRANULE is set, <b>-batch &lt;number&gt;</b> is also specified, the value from command line parameter <b>-batch</b> will be used. If neither the environment variable BATCH_SIZE_GRANULE is set nor <b>-batch</b> is specified, the user will be prompted to enter in runtime.                                                                                                                                                                                               |
| grbatch        | No        | The batch size used for physical granule file cleanup. If it is not provided in the command line, the user will be prompted to enter in runtime.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| phbatch        | No        | The <b>phbatch</b> size for PH granule deletion. Because PH granule deletion could be time consuming, set a high batch size for PH granule deletion could lock the database too long, so this parameter can be specified separately and keep a small value such as 5. This parameter can be omitted if the environment variable BATCH_SIZE_PH is set. If the environment variable BATCH_SIZE_PH is set, <b>-phbatch &lt;number&gt;</b> is also specified, the value from command line parameter <b>-phbatch</b> will be taken. If neither the environment variable BATCH_SIZE_PH is set nor <b>-phbatch</b> is specified, the user will be prompted to enter in runtime. |

#### 14.9.4.1 Deleting Granules, Phase 2: Running the Deletion Cleanup Utility

- 1 Log in at the x4dp101 host, where the Deletion Cleanup utility is installed.
- 2 To change to the directory for starting the Deletion Cleanup utility at the UNIX prompt enter:

```
cd /usr/ecs/<MODE>/CUSTOM/utilities
```

- The working directory is changed to /usr/ecs/<MODE>/CUSTOM/utilities.

3 To set up relevant environment variables (if desired) at the UNIX prompt enter:

```
setenv MODE <ECS mode>
setenv SYB_SQL_SERVER <xxdbl03_srvr>
setenv SYB_DB_NAME EcInDb_ <ECS mode>
```

4 To execute the Deletion Cleanup utility at the UNIX prompt enter:

```
EcDsDeletionCleanup.pl -user <db_user>-batch <number> -grbatch <number> -
phbatch <number> -log <log_file_name>
```

**NOTE:** Take care when increasing the BATCH\_SIZE\_GRANULE and BATCH\_SIZE\_PH values beyond the recommended default values. If the values are set too high, the database tables will be locked and all available Sybase locks will be used up.

- <batch size granule> represents the number of granules that will be deleted simultaneously from granule tables during granule cleanup. The default value is 50, which is accepted by pressing Return/Enter at the prompt without entering a value first.

5 Deletion Cleanup prepares to connect to the database and displays the following message is displayed:

**Ready to get into inventory database...**

6 After the Deletion Cleanup script connects to the database and checks for leftover granules that need to be processed. The following message is displayed:

*Previous run was not completed, you can choose to:*

1. *Rerun unfinished run only*
2. *Start a new run which include unfinished run(s)*
3. *Quit*

*Select 1, 2 or 3:*

7 Enter the appropriate number:

*<number>*

8 If there are granules marked as deletion or DFAed, the following message will be displayed for user selection:

*==== Menu for Lag Time ====*

1. *Select granules for a specific day (lag<n> or date <mm/dd/yyyy> format)*
2. *Select all granules older than a specific day (lag<n> or date <mm/dd/yyyy> format)*
3. *Quit*

*Select 1, 2 or 3:*

9 Enter the appropriate number:

*<number>*

- Entering 1 will cleanup granules whose deletion date fall into a single day specified by lag time. The user will be prompted to enter a lag time in number (*integer*) of days or a date **<mm/dd/yyyy>**.
- Entering 2 will cleanup all granules whose effective deletion date is older than the date specified by lag time. The user will be prompted to enter a lag time in number (*integer*) of days or a date **<mm/dd/yyyy>**.
- Entering 3 will cause the script to exit (i.e., nothing is cleaned up).

**10** After the lag time is confirmed, the utility will display the following menu for user selection:

**==== Menu for Data Type ====**

**1. Specify datatype(s) and version for deletion by an input file**

*The file format : one ESDT.Version <AST\_LIBT.001 or AST\_LIB\*.001> per line*

**2. Select all datatypes for deletion**

**3. Quit**

**Select 1, 2 or 3: \_**

**11** Enter 1, 2 or 3 as appropriate:

- Entering 1 will cause a subset of the listed ESDT.Version will be cleaned up by manual preparing an input file. Listing the selected ESDT.version in the file, one ESDT.version per line. A wildcard \* can be used for the ESDT only.
- Entering 2 will cleanup all ESDT.version.
- Entering 3 will cause the script to exit (i.e., nothing is cleaned up).
- The progress and failure information will be logged in the log file.

#### **14.9.4.2 Deleting Granules, Phase 2: Rerun unfinished Physical Cleanup**

---

The command for rerunning uncompleted run(s) is similar to starting a new run:

**►EcDsDeletionCleanup.pl -user <db\_user>**

**[-batch <number>]**

**[-grbatch <number>]**

**[-phbatch <number>]**

**[-log <log\_file\_name>]**

The utility always checks if there were any granule(s) left over from a previous unfinished run(s). If so, leftover information will be displayed and logged, a menu will be displayed for the user to select how to run the cleanup:

*Previous run was not completed, you can choose to:*

- 1. Rerun unfinished run only*
- 2. Start a new run which includes unfinished run(s)*
- 3. Quit*

*Select 1, 2 or 3:*

Select 1 to rerun unfinished run(s) will start to cleanup from the interrupted point, for instance, start to cleanup leftover XML files which had not been cleaned up in previous run(s).

Select 2 is the same as 14.6.4.1 but the cleanup will also include all leftovers in unfinished run(s).

---

## **14.10 Data Pool Maintenance Tasks**

### **14.10.1 Features of the Data Pool Maintenance GUI**

Most Archive or support personnel tasks for monitoring and maintaining the Data Pool require the use of the **Data Pool Maintenance (DPM) GUI**. The **DPM GUI** permits an operator to perform tasks in the following general areas:

- Monitoring Data Pool Active Insert Processes and Insert Actions.
- Managing Data Pool File Systems.
- Managing Cloud Cover Information.
- Checking the Status of Batch Inserts.
- Checking the Data Pool Insert Queue.
- Managing Data Pool Configuration Parameters and Data Pool Tuning.
- Managing Data Pool Collection Groups.
- Managing Data Pool Collections within Collection Groups.
- Managing Themes.

Other tasks are supported by scripts or utilities. For example, a Data Pool Update Expiration Script (Update Granule Utility) is available for extending the period of retention for selected science granules already in the Data Pool. There is a Data Pool cleanup utility that is typically run manually since on-line archive is deployed, but may be invoked through cron. Similarly, a utility for accumulating Data Pool access statistics is usually run in a cron job but may be invoked manually. There is a command line utility that permits operators to execute batch inserts of data from the archive into the Data Pool.

Distribution of data from the Data Pool is supported by the **HDF-EOS to GeoTIFF Conversion Tool (HEG)**. There are two versions of HEG:

- Data Pool HEG.
- Standalone HEG.

The Standalone HEG is a tool that an end user downloads and runs on his/her own workstation to convert EOS data products on the workstation from one format to another. The Data Pool HEG, which is accessed through the DAAC **Data Pool Web Access GUI** interface, is used to convert EOS data products before they are downloaded or shipped from the DAAC.

Finally, the **Spatial Subscription Server GUI** is a major Data Pool management tool. Although used primarily by User Services or science personnel, Archive or engineering support personnel may use it to extend the period of retention in a Data Pool insert subscription, and to view statistics on the processing of events and actions by the Spatial Subscription Server.

Both the **DPM GUI** and the **Spatial Subscription Server GUI** are web applications, accessed through the Mozilla Firefox 2.0 standard web browser.

Table 14.10-1 provides an Activity Checklist for Data Pool Maintenance Tasks addressed in this section.

**Table 14.10-1. Data Pool Maintenance Tasks - Activity Checklist (1 of 2)**

| Order | Role               | Task                                                                  | Section       | Complete? |
|-------|--------------------|-----------------------------------------------------------------------|---------------|-----------|
| 1     | Archive Technician | Launch the DPM GUI                                                    | (P) 14.10.1.1 |           |
| 2     | Archive Technician | Shut Down the DPM GUI                                                 | (P) 14.10.1.2 |           |
| 3     | Archive Technician | Monitor Data Pool Active Insert Processes                             | (P) 14.10.1.3 |           |
| 4     | Archive Technician | View a List of Data Pool File Systems                                 | (P) 14.10.2.1 |           |
| 5     | Archive Technician | Add a Data Pool File System                                           | (P) 14.10.2.2 |           |
| 6     | Archive Technician | Modify a Data Pool File System                                        | (P) 14.10.2.3 |           |
| 7     | Archive Technician | View cloud Cover Information                                          | (P) 14.10.3.1 |           |
| 8     | Archive Technician | Add New Cloud Cover Information                                       | (P) 14.10.3.2 |           |
| 9     | Archive Technician | Modify Cloud Cover Information                                        | (P) 14.10.3.3 |           |
| 10    | Archive Technician | Delete Cloud Cover Information                                        | (P) 14.10.3.4 |           |
| 11    | Archive Technician | Check the Status of Batch Inserts                                     | (P) 14.10.4.1 |           |
| 12    | Archive Technician | Check the Data Pool Insert Queue and Cancel a Data Pool Insert Action | (P) 14.10.5.1 |           |

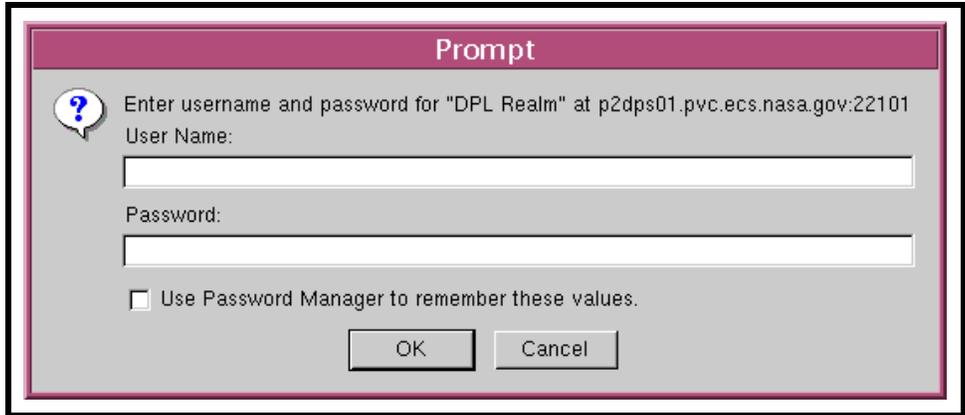
**Table 14.10-1. Data Pool Maintenance Tasks - Activity Checklist (2 of 2)**

| Order | Role               | Task                                        | Section       | Complete? |
|-------|--------------------|---------------------------------------------|---------------|-----------|
| 13    | Archive Technician | View DPM Configuration Parameter Values     | (P) 14.10.6.1 |           |
| 14    | Archive Technician | Modify DPM Configuration Parameter Values   | (P) 14.10.6.2 |           |
| 15    | Archive Technician | View DPM Aging Parameter Values             | (P) 14.10.7.1 |           |
| 16    | Archive Technician | Modify DPM Aging Parameter Values           | (P) 14.10.7.2 |           |
| 17    | Archive Technician | View Collection Groups                      | (P) 14.10.8.1 |           |
| 18    | Archive Technician | Modify Collection Groups                    | (P) 14.10.8.2 |           |
| 19    | Archive Technician | Add a Collection Group                      | (P) 14.10.8.3 |           |
| 20    | Archive Technician | Delete a Collection Group                   | (P) 14.10.8.4 |           |
| 20    | Archive Technician | Add an ECS Collection to a Collection Group | (P) 14.10.8.5 |           |
| 21    | Archive Technician | Modify an ECS Collection                    | (P) 14.10.8.6 |           |
| 22    | Archive Technician | View a List of Themes                       | (P) 14.10.9.1 |           |
| 23    | Archive Technician | Filter a List of Themes                     | (P) 14.10.9.2 |           |
| 24    | Archive Technician | Modify a Theme                              | (P) 14.10.9.3 |           |
| 25    | Archive Technician | Add a Theme                                 | (P) 14.10.9.4 |           |
| 26    | Archive Technician | Delete a Theme                              | (P) 14.10.9.5 |           |

Let's examine how the **DPM GUI** is used for Data Pool maintenance tasks. Of course, the first thing to do is launch the GUI. The procedure for launching the GUI is provided separately here and is referenced in other procedures. It applies to both full-capability and limited-capability operators. An operator's level of permission is determined when the operator logs in to the GUI using the security login prompt (Figure 14.10-1).

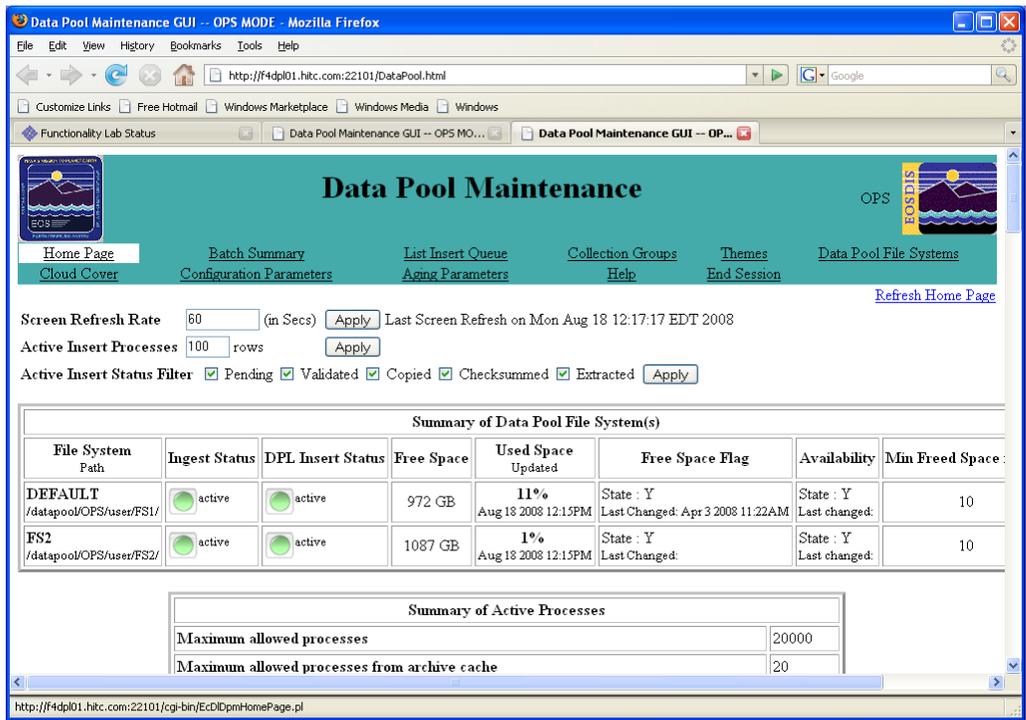
Full-capability operators have the ability to configure parameters and perform all other actions that can be accomplished with the GUIs. Limited-capability operators are able to view a lot of information; however, on the limited-capability GUI some buttons and links have been disabled so it is not possible to perform certain actions or access certain pages.

This lesson provides instruction in the full-capability version of the GUIs. However, the functions that are available to limited-capability operators as well as the functions that are not available to limited-capability operators are identified.



**Figure 14.10-1. Security Login Prompt**

Figure 14.10-2 illustrates the **DPM GUI Home Page**, from which the operator can perform some monitoring and maintenance tasks and from which there is access to other pages supporting other tasks.



**Figure 14.10-2. DPM GUI Home Page**

The **DPM GUI Home Page** (Figure 14.10-2) displays the state of several parameters and allows an operator to make changes. It also lists active insert processes. Near the top of the **Home Page** are links allowing an operator to access other functions including the following items:

- **Data Pool File Systems**
- **Cloud Cover**
- **List Insert Queue**
- **Batch Summary**
- **Collection Groups**
- **Themes**
- **Configuration Parameters**
- **Aging Parameters**

There is also a **Help** page for assistance in navigation of the GUI and an **End Session** link for logging out of the GUI.

#### 14.10.1.1 Launch the DPM GUI

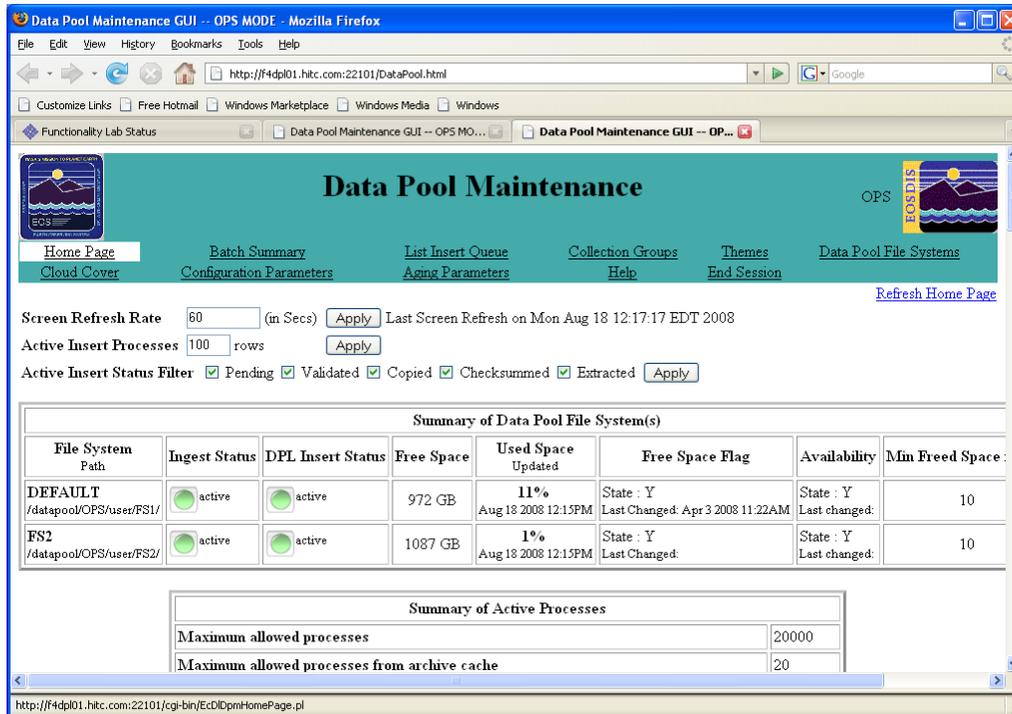
---

- 1 At the UNIX command shell prompt, type **setenv DISPLAY *clientname*:0.0** and then press the **Return/Enter** key.
  - For *clientname*, use either the local terminal/workstation IP address or its machine name.
- 2 Start the log-in to a Firefox host by typing **/tools/bin/ssh *hostname*** (x4dpl01) at the UNIX command shell prompt, and press the **Return/Enter** key.
  - If you receive the message, **Host key not found from the list of known hosts. Are you sure you want to continue connecting (yes/no)?** type **yes** (“y” alone does not work).
  - If you have previously set up a secure shell passphrase and executed **sshremote**, a prompt to **Enter passphrase for RSA key '<user@localhost>'** appears; continue with Step 3.
  - If you have not previously set up a secure shell passphrase, go to Step 4.
- 3 If a prompt to **Enter passphrase for RSA key '<user@localhost>'** appears, type your **Passphrase** and then press the **Return/Enter** key. Go to Step 5.
- 4 At the **<user@remotehost>'s password:** prompt, type your **Password** and then press the **Return/Enter** key.
  - You are logged in and a UNIX command shell prompt is displayed.

- 5 Type **Firefox &** then press **Return/Enter**.
  - It may be necessary to type the path as well as the Firefox command (e.g., **/tools/bin/Firefox &**).
  - It may be necessary to respond to dialogue boxes, especially if the browser is already being used by someone else who has logged in with the same user ID.
  - The Firefox web browser is displayed.
- 6 If a bookmark has been created for the DPM GUI, select the appropriate bookmark from those listed on the browser.
  - The security login **Prompt** is displayed.
- 7 If no bookmark has been created for the **DPM GUI**, type **http://host:port/path** in the browser's **Location (Go To)** field then press **Return/Enter**.
  - For example: **http://x4dpl01.daac.ecs.nasa.gov:54321/DataPool.html**
  - The security login **Prompt** is displayed.
- 8 Type the appropriate user name in the **User Name** box of the security login **Prompt**.
- 9 Type the appropriate password in the **Password** box of the security login **Prompt**.

**NOTE:** If the security login prompt reappears after the first time the user name and password have been entered (and the **OK** button has been clicked), it may not be due to a data entry problem. Try again to log in using the same user name and password. Sometimes it is necessary to enter the user name and password for the GUI more than once.

- 10 Click on the appropriate button from the following selections:
  - **OK** - to complete the log-in and dismiss the dialogue box.
    - The dialogue box is dismissed.
    - The **DPM GUI Home Page** is displayed (see Figure 14.10-3).



**Figure 14.10-3. DPM GUI Home Page**

- **Cancel** - to dismiss the dialogue box without logging in.
  - The dialogue box is dismissed.
  - The Firefox web browser is displayed.

At some point it becomes necessary to shut down the **DPM GUI** (end a **DPM GUI** session). The procedure that follows is recommended and is applicable to both full-capability and limited-capability operators.

#### 14.10.1.2 Shut Down the DPM GUI (End a DPM GUI Session)

- 1 Click on the **Home Page** link at the top of the **DPM GUI**.
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **End Session** link at the top of the **Home Page**.
  - A log-out page containing the following message is displayed.

**“Click on Button Below to End Session:**

**THIS WOULD ALSO SHUT DOWN THE BROWSER:”**

**NOTE:** To abort the log-out and return to the **Home Page**, click on the browser **Back** button.

- 3 Click on the **ShutDown** button.
    - The Firefox browser is dismissed.
- 

You may wish to keep an instance of the **DPM GUI** displayed to monitor Data Pool Active Insert Processes. The procedure for using the **DPM GUI** to monitor Data Pool active insert processes is applicable to both full-capability and limited-capability operators.

#### 14.10.1.3 Monitor Data Pool Active Insert Processes

---

- 1 Launch the **DPM GUI**.
  - The **Home Page** is displayed.
- 2 Observe information displayed on the **DPM GUI Home Page**.
  - The **Home Page** has the following links for access to Data Pool maintenance function pages:
    - **Data Pool File Systems**
    - **Cloud Cover**
    - **List Insert Queue**
    - **Batch Summary**
    - **Collection Groups**
    - **Themes**
    - **Configuration Parameters**
    - **Aging Parameters**
    - **End Session**
  - The **Home Page** has a **Summary of Data Pool File Systems** with the following columns:
    - **File System** (representing an existing Data Pool file system).
    - **Ingest Status**
    - **DPL Insert Status**
    - **Free Space**
    - **Used Space**
    - **Free Space Flag** (if set to “Y,” free space is available for inserts; “N” means free space is not available).

- **Availability** (if set to “Y,” the file system is currently available for Data Pool insert; “N” means the file system is not available for Data Pool insert).
- **Min Freed Space in MB** (value that represents the minimum amount of freed space in the file system in megabytes; it is an amount of space must remain free in order to make the file system available for insert).
- The **Home Page** has a **Summary of Active Processes** with the following rows:
  - **Maximum allowed processes**
  - **Maximum allowed processes from archive cache**
  - **Maximum allowed processes from archive tape**
  - **Total number of active insert processes running**
  - **Total number of validated active insert processes running**
  - **Total number of pending active insert processes running**
  - **Number of active insert processes using archive cache**
  - **Number of active insert processes using archive tape**
- The **Home Page** has a table of **List of Active Insert Processes ( Rows X)** showing the following columns of detailed information for each process:
  - **Unix ProcessId** (UNIX process identifier).
  - **EcsID** (ECS identifier or Granule ID for the granule being processed).
  - **Collection** (to which the granule belongs).
  - **Version** (for the collection to which the granule belongs).
  - **StartTime** (time at which the insert processing started).
  - **StatusTime** (time at which the status listed in the Status column was achieved).
  - **Status** (current state of the insert process).
  - **Archive Cache** [availability (Y or N) of the granule being processed].
  - **Retries** [number of attempts by the process to recover from retryable errors (e.g., Data Pool disk temporarily unavailable, Data Pool directory does not exist, or Data Pool database temporarily unavailable)].

**NOTE:** The system is designed for rapid insertion of data into the Data Pool by quickly processing data that are available in cache, such as data that are staged for archiving. If the insert processing is delayed and the data are removed from cache, the Data Pool insert is likely to fail.

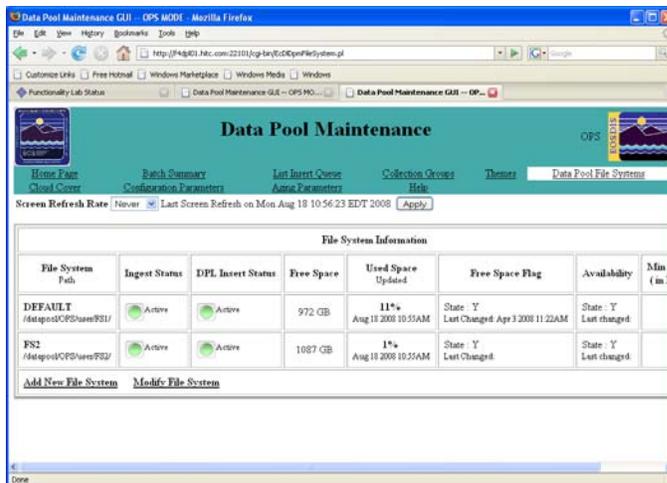
- 3 To obtain an immediate screen refresh, click on the **Refresh Home Page** link near the upper right corner of the display.
  - The displayed data are updated.

**NOTE:** The screen refreshes automatically at intervals determined by the number of seconds specified in the **Screen Refresh Rate** field.

- 4 To change the automatic screen refresh rate first type the desired number of seconds between refreshes in the **Screen Refresh Rate** text entry box.
  - 5 To complete changing the automatic screen refresh rate click on the **Apply** button adjacent to the **Screen Refresh Rate** text entry box.
    - The **Screen Refresh Rate** is changed to the new value.
  - 6 To change the number of active insert processes displayed at a time in the **List of Active Insert Processes** table on the **Home Page** first type the desired number of rows to be displayed in the **Active Insert Processes** text entry box.
  - 7 To complete changing the number of active insert processes displayed at a time in the **List of Active Insert Processes** table on the **Home Page** click on the **Apply** button adjacent to the **Active Insert Processes** text entry box.
    - The number of active insert processes displayed at a time in the **List of Active Insert Processes** table is changed to the new value.
  - 8 To filter the list of Active Insert processes in the **List of Active Insert Processes** table on the **Home Page** based on the Active Insert's status. Click the checkbox for any combination of the five possible filters.
  - 9 To complete filter(s) selection click on the **Apply** button adjacent to the **Active Insert Status Filter** checkboxes.
- 

### 14.10.2 Data Pool File Systems

Figure 14.10-4 illustrates the Data Pool **File System Information** page that allows both full-capability and limited-capability operators to view a list of Data Pool file systems and obtain information on **Free Space Flag**, **Availability** for insert, and **Minimum Freed Space**. From this page, the full-capability operator is able to configure a new file system or modify an existing file system (which may include assigning Availability and/or No Free Space status).



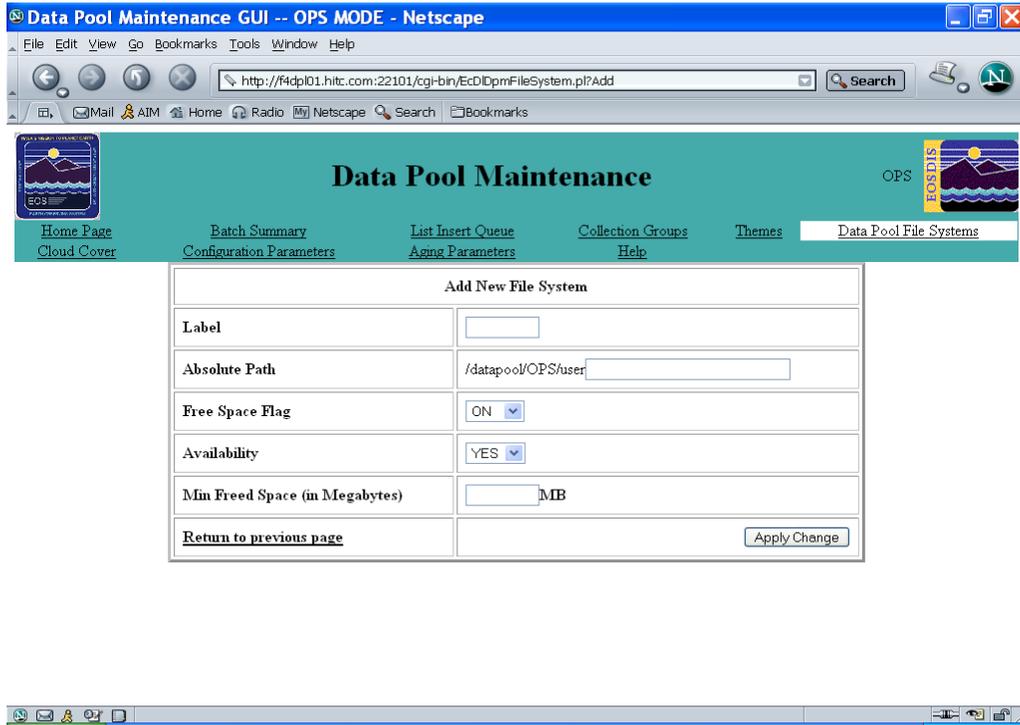
**Figure 14.10-4. Data Pool File System Page**

### 14.10.2.1 View a List of Data Pool File Systems

- 1 Launch the **DPM GUI**.
  - The **DPM GUI Home page** is displayed.
- 2 Click on the **Data Pool File Systems** link.
  - The **File System Information** page is displayed (see Figure 14.10-4).
- 3 Observe data displayed on the **File System Information** page.
  - The table on the File System Information page has columns containing the following types of Data Pool file system information:
    - **File System Path**
    - **Ingest Status**
    - **DPL Inset Status**
    - **Free Space**
    - **Used Space Updated**
    - **Free Space Flag**
    - **Availability**
    - **Min Freed Space (in Megabytes)**
  - The following links are available on the **File System Information** page:
    - **Add New File System**
    - **Modify File System**

4 Clicking on the **Add New File System** link takes the full-capability operator to the **Add New File System** page shown in Figure 14.10-5. The operator needs to add data in the following five fields:

- [File System] **Label**: a label representing an existing Data Pool file system;
- **Absolute Path**: the path to the directory where the file system is located (the basic ftp root path is provided and the operator completes the path name if necessary);
- **Free Space Flag**: a value that needs to be set to either “ON” or “OFF” (ON means free space is available for inserts; OFF means free space is not available);
- **Availability**: a value that needs to be set to either “YES” or “NO” (YES means the file system is currently available for Data Pool insert; NO means the file system is not available for Data Pool insert);
- **Min Freed Space (in Megabytes)**: an integer value that represents the minimum amount of freed space in the file system in megabytes; it is an amount of space must remain free in order to make the file system available for insert.



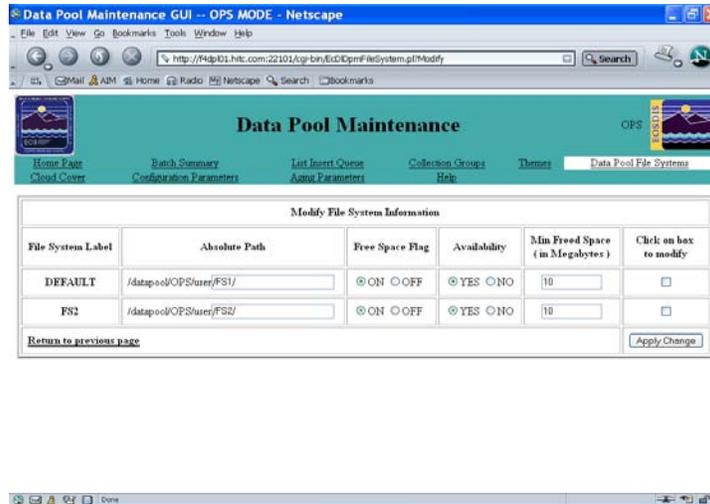
**Figure 14.10-5. Add New File System Page**

### 14.10.2.2 Add a Data Pool File System

---

- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home page** is displayed.
  - 2 Click on the **Data Pool File Systems** link.
    - The **File System Information** page is displayed.
  - 3 Click on the **Add New File System** link at the bottom of the list of file systems (scrolling down if necessary).
    - The **Add New File System** page is displayed (see Figure 14.10-5).
  - 4 Type the desired file system label in the **Label** field.
    - Enter a unique name with no more than 25 characters.
  - 5 Type the desired path in the **Absolute Path** field.
    - The basic ftp root directory path is shown adjacent to the text entry box; data entered in the box will be appended to the base path shown.
  - 6 To display free space flag options click on the **Free Space Flag** option button.
    - **Free Space Flag** options are displayed (i.e., **ON** and **OFF**).
  - 7 To display availability options click on the **Availability** option button.
    - **Availability** options are displayed (i.e., **YES** and **NO**).
  - 8 To select an availability option click on the appropriate choice from the option list.
    - **YES** should be selected if the file system is currently available for inserts.
    - **NO** should be selected if the file system is not currently available for inserts.
  - 9 Type the desired value for minimum freed space (in megabytes) in the Min Freed Space (in Megabytes) field.
    - **Min Freed Space** indicates how much space needs to be available to keep the file system available for insert.
  - 10 Click on the **Apply Change** button.
    - The file system information is entered in the Data Pool database.
    - The **File System Information** page is displayed with the new file system information.
- 

The **DPM GUI** may be used to modify a Data Pool file system. This is useful if the Absolute Path, Free Space Flag, Availability (for Insert), and/or Min. Freed Space for a particular Data Pool file system need to be corrected or updated. Selecting the **Modify File System** link takes the full-capability operator to the **Modify File System Information** page shown in Figure 14.10-6. The operator can change the Absolute Path, Free Space Flag, Availability flag, or the Min Freed Space on this page. There are check boxes associated with each file system. The operator can change multiple file systems at one time by checking the desired file systems' checkboxes and clicking on the **Apply Change** button.



**Figure 14.10-6. Modify File System Information Page**

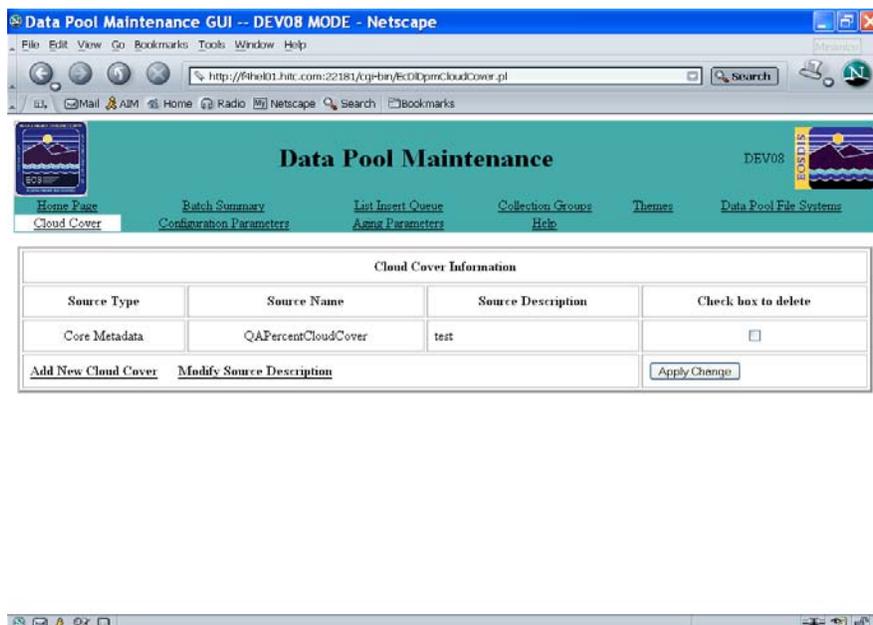
### 14.10.2.3 Modify a Data Pool File System

- 1 Launch the **DPM GUI**.
  - For detailed instructions refer to the Launch the DPM GUI procedure (previous section of this lesson).
  - The **DPM GUI Home page** is displayed.
- 2 Click on the **Data Pool File Systems** link.
  - The **File System Information** page is displayed.
- 3 Click on the **Modify File System** link at the bottom of the list of file systems (scrolling down if necessary).
  - The **Modify File System Information** page is displayed (see Figure 14.10-6).
- 4 To change the absolute path for a file system type the desired path in the **Absolute Path** field for the file system.
  - The basic ftp root directory path is shown above the text entry box; data entered in the box will be appended to the base path shown.
- 5 To change a file system's free space flag setting click on the appropriate button in the **Free Space Flag** column.
  - The following choices are available:
    - **ON**
    - **OFF**
- 6 To change the setting for a file system's availability for data insert click on the appropriate button in the **Availability** column.

- The following choices are available:
    - YES
    - NO
- 7 To change the minimum freed space for a file system type the desired value (in megabytes) in the appropriate **Min Free Space (in Megabytes)** field.
  - 8 Click in the check box at the end of the row containing file system information to be modified.
    - The selected file system information is marked for subsequent modification.
  - 9 Repeat Steps 4 through 8 for any additional file systems to be modified.
  - 10 Click on the **Apply Change** button.
    - The revised file system information is entered in the Data Pool database.
    - The **File System Information** page is displayed with the modified file system information.
- 

### 14.10.3 Cloud Cover

Both full-capability and limited-capability operators can view existing cloud cover information in the Data Pool database by clicking on the **Cloud Cover** link shown in Figure 14.10-2. The link takes the operator to the **Cloud Cover Information** page shown in Figure 14.10-7. The page displays the information concerning the sources of cloud cover; i.e., the **Source Type**, **Source Name**, and **Source Description**.



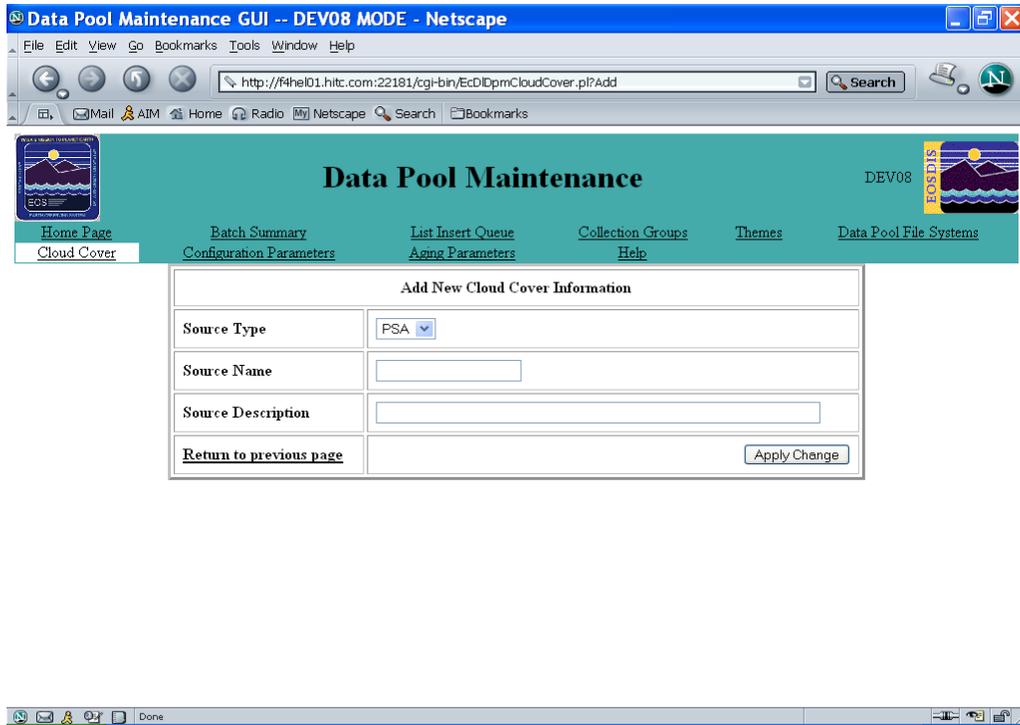
**Figure 14.10-7. Cloud Cover Information Page**

### 14.10.3.1 View Cloud Cover Information

---

- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home page** is displayed.
  - 2 Click on the **Cloud Cover** link.
    - The **Cloud Cover Information** page is displayed (see Figure 14.10-7).
  - 3 Observe data displayed on the **Cloud Cover Information** page.
    - The table on the **Cloud Cover Information** page has columns containing the following types of cloud cover information:
      - **Source Type**
      - **Source Name**
      - **Source Description**
      - **Check box to delete**
    - The following links are available on the **Cloud Cover Information** page:
      - **Add New Cloud Cover**
      - **Modify Source Description**
    - An **Apply Change** button is available for deleting cloud cover information from the Data Pool database.
- 

The full-capability operator can add a new cloud cover source by clicking on the **Add New Cloud Cover** link shown in Figure 14.10-7. The link takes the operator to the **Add New Cloud Cover Information** page shown in Figure 14.10-8. After selecting the Source Type from an option list and entering the Source Name and Source Description, the operator clicks on the **Apply Change** button. All Source Names are validated against the Inventory database. The new cloud cover source is added to the Data Pool database and the **Cloud Cover Information** page is refreshed.



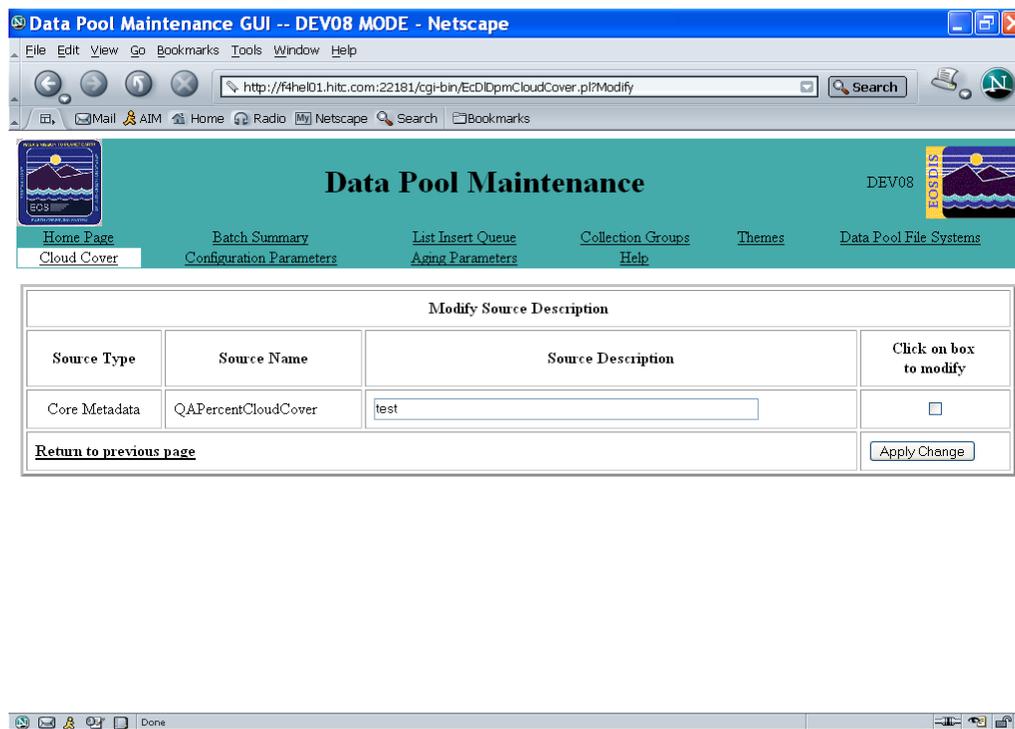
**Figure 14.10-8. Add New Cloud Cover Information Page**

### 14.10.3.2 Add New Cloud Cover Information

- 1 Launch the **DPM GUI**
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Cloud Cover** link.
  - The **Cloud Cover Information** page is displayed.
- 3 Click on the **Add New Cloud Cover** link at the bottom of the **Cloud Cover Information** page (scrolling down if necessary).
  - The **Add New Cloud Cover Information** page is displayed (see Figure 14.10-8).
- 4 To view the source type options click on the **Source Type** option button.
  - Source type options are displayed (e.g., **Core Metadata** and **PSA**).
- 5 To select a source type click on the appropriate source type from the option list.
  - If **Core Metadata** was selected, the **Source Name** field is automatically filled in.
- 6 To specify a source name type the desired name in the **Source Name** field.
  - If **Core Metadata** was selected as the source type, the **Source Name** field is automatically filled in and cannot be edited.

- 7 Type a description of the cloud cover information in the **Source Description** field.
    - The description may be up to 255 characters in length.
  - 8 Click on the **Apply Changes** button.
    - The source name is validated against the Inventory database.
    - The new cloud cover information is entered in the Data Pool database.
    - The **Cloud Cover Information** page is displayed with the new cloud cover information.
- 

The full-capability operator can modify an existing cloud cover Source Description by clicking on the **Modify Source Description** link shown in Figure 14.10-7. The link takes the operator to the **Modify Source Description** page shown in Figure 14.10-9. The operator can modify the Source Descriptions only. (To modify a Source Type or Source Name the operator must delete the applicable cloud cover information row and add a new one with the correct information.) After making desired changes, the operator clicks on the checkbox(es) adjacent to the source(s) to be modified and clicks on the **Apply Change** button. The changes are applied to the Data Pool database and the **Cloud Cover Information** page is refreshed.



**Figure 14.10-9. Modify Source Description Page**

### 14.10.3.3 Modify Cloud Cover Information

---

- 1 Launch the **DPM GUI**
    - For detailed instructions refer to the **Launch the DPM GUI** procedure (previous section of this lesson).
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Cloud Cover** link.
    - The **Cloud Cover Information** page is displayed.
  - 3 Click on the **Modify Source Description** link at the bottom of the **Cloud Cover Information** page (scrolling down if necessary).
    - The **Modify Source Description** page is displayed (see Figure 14.10-9).
  - 4 To start the process of changing a source description type the desired description in the appropriate **Source Description** field.
  - 5 Click in the check box at the end of the row containing modified source description information.
    - The source description is marked for subsequent modification. (A check mark is displayed in the selected check box.)
  - 6 Repeat Steps 4 and 5 for any additional source descriptions to be modified.
  - 7 Click on the **Apply Change** button.
    - The revised source description information is entered in the Data Pool database.
    - The **Cloud Cover Information** page is displayed with the modified cloud cover information.
- 

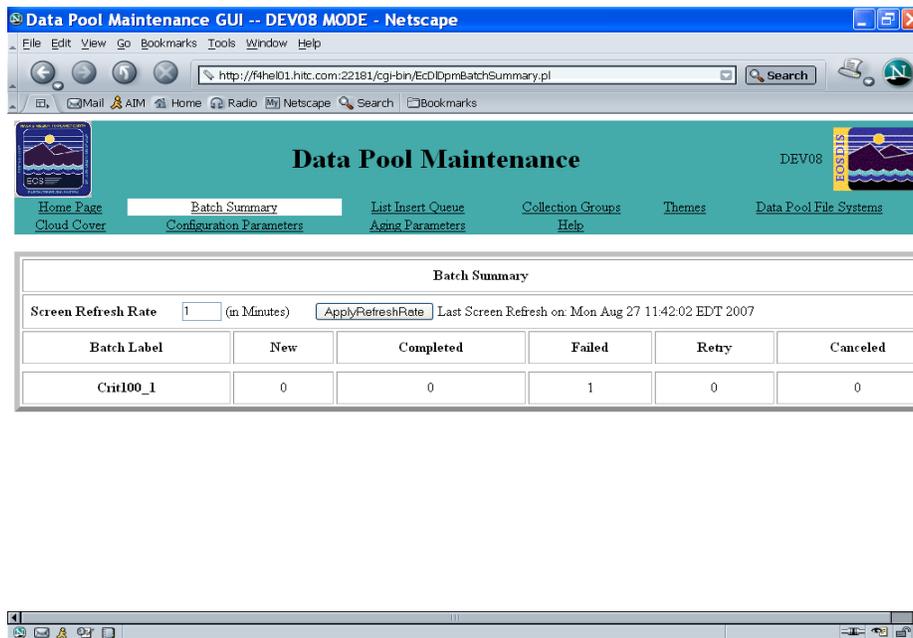
### 14.10.3.4 Delete Cloud Cover Information

---

- 1 Launch the **DPM GUI**
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Cloud Cover** link.
    - The **Cloud Cover Information** page is displayed.
  - 3 Click in the check box(es) at the end of the row(s) containing the **Check box to delete**.
    - The selected source(s) is (are) marked for subsequent deletion.
  - 4 Click on the **Apply Change** button.
    - The selected source(s) is (are) deleted from the Data Pool database.
    - If any cloud cover information is associated with any collection, it will not be deleted.
    - The **Cloud Cover Information** page is displayed with the modified cloud cover information.
-

## 14.10.4 Batch Summary

Figure 14.10-10 illustrates the **Batch Summary** page, which is accessible from the **Batch Summary** link on the **DPM GUI Home Page** (Figure 14.10-2). The **Batch Summary** page displays information on inserts made with the command line utility that permits operators to execute batch inserts of data from the archive into the Data Pool. In addition, it displays a summary of the status of Data Pool inserts for each batch label. Insert statuses include “new,” “completed,” “failed,” “retry,” and “canceled.” The information is accessible to both full-capability and limited-capability operators.



**Figure 14.10-10. Batch Summary Page**

The **DPM GUI** provides a page to display a summary of the status of batch Data Pool inserts made using the batch insert utility. The procedure that follows is applicable to both full-capability and limited-capability operators.

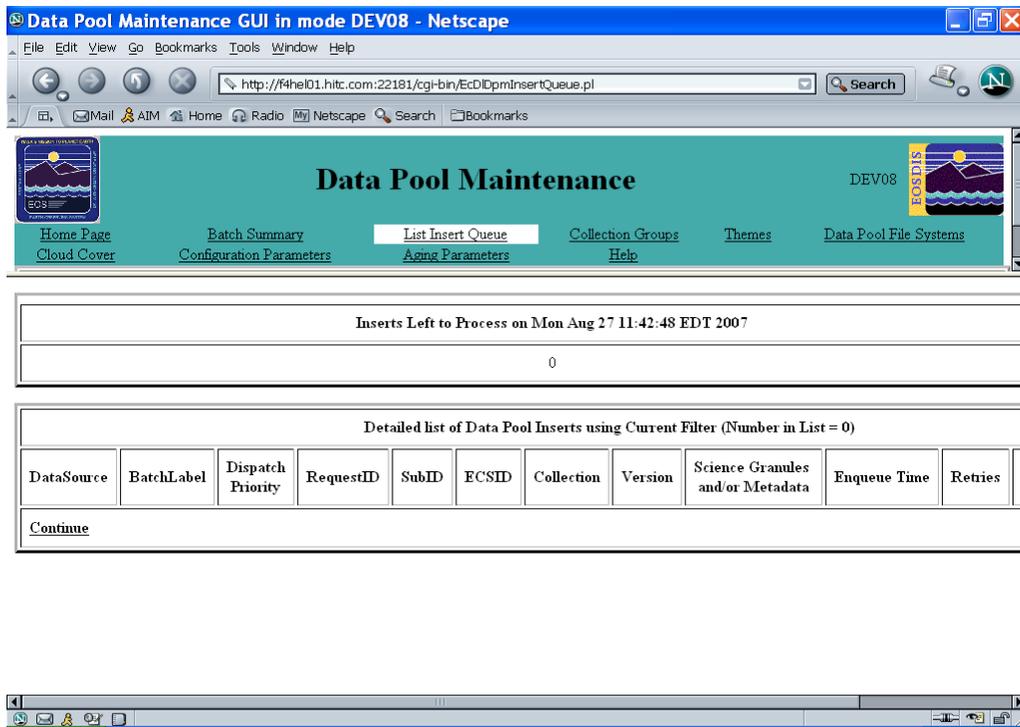
### 14.10.4.1 Check the Status of Batch Inserts

- 1 Launch the **DPM GUI**
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Batch Summary** link.
  - The **Batch Summary** page is displayed (see Figure 14.10-10).

- 3 Observe data displayed on the **Batch Summary** page.
    - The table on the **Batch Summary** page has columns containing the following types of information:
      - **Batch Label**
      - **New** (number of inserts for the label that are new)
      - **Completed** (number of inserts for the label that have been completed)
      - **Failed** (number of inserts for the label that have failed)
      - **Retry** (number of inserts for the label that have been retried)
      - **Canceled** (number of inserts for the label that have been canceled)
  - 4 To change the automatic screen refresh rate first type the desired number of minutes between refreshes in the **Screen Refresh Rate** text entry box.
  - 5 To complete changing the automatic screen refresh rate click on the **ApplyRefreshRate** button adjacent to the **Screen Refresh Rate** text entry box.
    - The **Screen Refresh Rate** is changed to the new value.
  - 6 Return to Step 3.
- 

#### 14.10.5 List Insert Queue

Figure 14.10-11 illustrates the **List Insert Queue** page, which is accessible from the **List Insert Queue** link on the **DPM GUI Home Page** (Figure 14.10-2). The page provides a list with detailed information on inserts left to process. The information is accessible to both full-capability and limited-capability operators.



**Figure 14.10-11. List Insert Queue Page**

The **List Insert Queue** page of the **DPM GUI** provides a list of Data Pool inserts left to process that both full-capability and limited-capability operators can view. It also provides for each listed insert a check box permitting a full-capability operator to mark queued inserts for cancellation, and an **Apply Change** button to implement the cancellation.

#### **14.10.5.1 Check the Data Pool Insert Queue and Cancel a Data Pool Insert Action**

- 1** Launch the **DPM GUI**
  - The **DPM GUI Home Page** is displayed.
- 2** Click on the **List Insert Queue** link.
  - The **List Insert Queue** page is displayed (see Figure 14.10-11).
- 3** Observe data displayed on the **List Insert Queue** page.
  - The **List Insert Queue** page shows how many inserts are left to process as of the current date.
  - The table on the **List Insert Queue** page has columns containing the following types of insert queue information:
    - **Data Source**

- **Batch Label**
- **Dispatch Priority**
- **RequestID**
- **SubID** (subscription identifier of the subscription selected by the software for processing)
- **ECSID** (ECS identifier or Granule ID for the granule to be processed)
- **Collection** (to which the granule belongs)
- **Version** (for the collection to which the granule belongs)
- **Science Granules and/or Metadata** (indication of whether the insert is to include science granules and metadata or just the metadata)
- **Enqueue Time** (time when the insert was placed in the insert queue)
- **Retries** [number of attempts by the process to recover from retryable errors (e.g., Data Pool disk temporarily unavailable, Data Pool directory does not exist, Data Pool database temporarily unavailable)]
- **Status**
- **Click on Box to Cancel** (containing a check box to mark the insert for cancellation)

**NOTE:** There may be multiple subscriptions specifying insertion of specific data into the Data Pool, but only one insert is needed; therefore, only one of the subscriptions serves as the basis for the insert action. The **SubID** is of no particular significance to an operator and may safely be ignored.

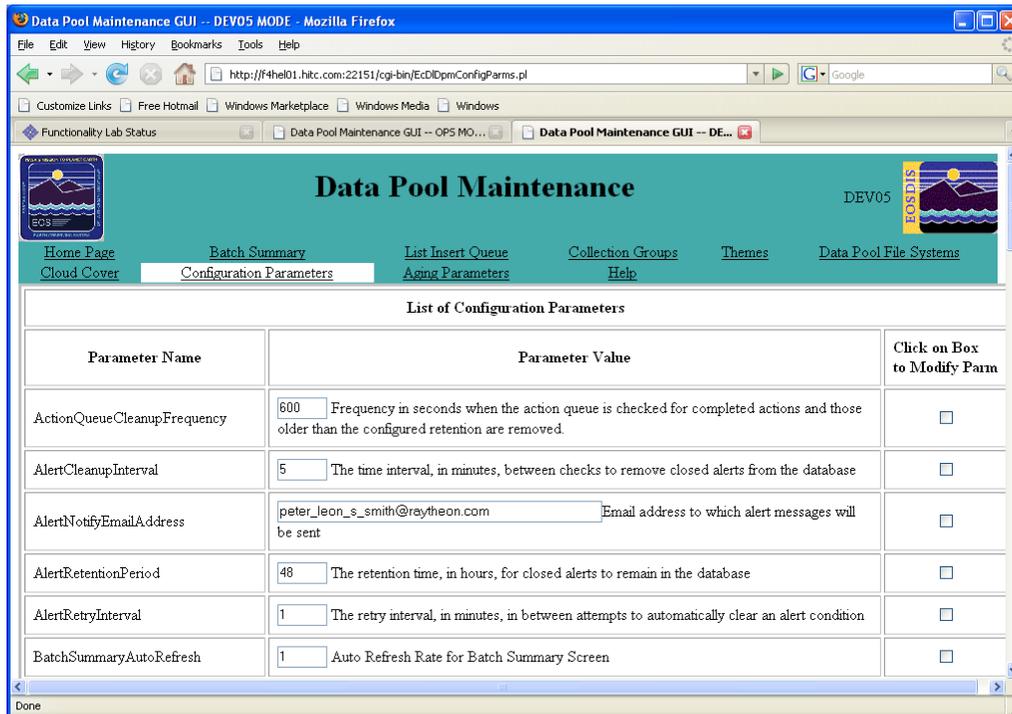
- There is an **Apply Change** button at the bottom of the page for implementing cancellations.
  - There is a **Continue** link at the bottom of the page; if there are more inserts than can be displayed in the space of one page, the **Continue** link displays the next page of the list.
- 4** To cancel an insert, first click on the check box at the end of the row of information for the insert to be canceled.
- The insert is marked for subsequent cancellation.
  - The check box for the selected insert is filled to indicate selection.
- 5** Repeat Step 4 for any additional insert to be canceled.
- 6** To implement the cancellation of insert(s) click on the **Apply Change** button.
- A confirmation message is displayed; it asks "**Are you ready to cancel the insert for . . .**" and there are links displayed for **Yes, cancel insert** and **No**, return to previous page.

- 7 To confirm cancellation, click on the **Yes, cancel insert** link.
- The **List Insert Queue** page is displayed with the canceled insert(s) removed and the count of inserts left to process reduced by the number of inserts canceled.

### 14.10.6 Configuration Parameters

Figure 14.10-12 shows the **List of Configuration Parameters** page, which is accessible from the **Configuration Parameters** link on the **DPM GUI Home Page** (Figure 14.10-2). The page lists numerous Data Pool configuration parameters with their settings and a brief description of each. The information is accessible to both full-capability and limited-capability operators.

For each parameter there is a text box or option list in the **Parameter Value** column so the full-capability operator can assign a new value to the parameter when necessary. In addition, there is a check box that the full-capability operator uses to mark parameters with values to be modified. At the bottom of the page is an **Apply Change** button for implementing the change(s).



**Figure 14.10-12. List of Configuration Parameters Page**

The following parameters are examples of the types of parameters in the Data Pool database that the full-capability operator can modify:

- ActionQueueCleanupFrequency - frequency in seconds when the action queue is checked for completed actions and those older than the configured retention period are removed.
- AlertCleanupInterval - The time interval, in minutes, between checks to remove closed alerts from the database.
- AlertNotifyEmail Address - Email address to which alert messages will be sent.
- AlertRetentionPeriod - The retention time, in hours, for closed alerts to remain in the database.
- AlertRetryInterval - The retry interval, in minutes, in between attempts to automatically clear an alert condition.
- BatchSummaryAutoRefresh – autorefresh rate for the Batch Summary page.
- ChecksumPercent - The percentage of science files that will have checksums verified or computed.
- Clean703Orders -Flag indicating whether DPL should clean up order only granules.
- DPLRetentionPatchInstalled - The existence of this configuration parameter means that the DPL Retention patch has been installed and granules will not expire from the Data Pool.
- DatabaseRetryCount - The number of times a retryable database error may be retried before being considered failed.
- Database RetryInterval - The number of seconds to wait between retries of a retryable database error.
- DefaultRetentionPeriod - default retention period in days for all Data Pool Insert Actions.
- DefaultRetentionPriority - default retention priority for all Data Pool Inserts actions. The range of valid values is 1 – 255.
- DeleteCompletedActionsAfter - time in minutes that operators let completed actions stay in the insert action queue before making them eligible for removal. The delay is intended to provide the operator with some ability to check on past actions. The time period should not be too long.
- DisplayAIPChunkSize - number of rows to return per chunk for the Active Insert Processes list.
- FileSystemCheckInterval - The time interval, from 1 to 10 minutes, in between attempts to automatically clear a Data Pool file system alert condition.
- FileSystemRefreshRate - Time in minutes before the File Systems Page Refreshes.

- FilterChecksumAIP - Show Checksummed Active Insert Processes on the Data Pool Maint. GUI page.
- FilterCopiedAIP - Show Copied Active Insert Processes on the Data Pool Maint. GUI page.
- FilterExtractedAIP - Show Extracted Active Insert Processes on the Data Pool Maint. GUI page.
- FilterPendingAIP - Show Pending Active Insert Processes on the Data Pool Maint. GUI page.
- FilterValidAIP - Show Validated Active Insert Processes on the Data Pool Maint. GUI page.
- FreeSpaceResumePercent - The percentage of free space required before a Data Pool file system full condition may be cleared.
- GranuleLockRetentionPeriod - The age in hours that determines when a granule lock should be considered stale
- GranuleOmLockRetentionPeriod - The age in minutes that determines when a granule lock by OMS should be considered stale
- HEGCleanupAge – HDF-EOS to GeoTIF Converter (HEG) cleanup age in days.
- IdleSleep - number of seconds to sleep when there is nothing to do (Obsolete in 7.20).
- InCacheTimeLimit - maximum time in minutes that operators are willing to wait for a Data Pool Insert Utility (DPIU) process to complete when its files are in cache. When the time limit is reached, the Data Pool Action Driver (DPAD) kills the process and retries the action (Obsolete in 7.20)
- InsertRetryWait - number of seconds to wait before an insert that failed should be resubmitted (if it can be retried).
- MAX\_READ\_DRIVES\_x0xxgnn – (multiple parameters as necessary) maximum number of simultaneous tape drives used for the specified archive.
- MFSONInsert – specifies whether or not (YES or NO) DPAD should use the Multiple File System table (Obsolete in 7.20).
- MaxConcurrentBandExtract – The maximum number of concurrent Band Extraction operations.
- MaxConcurrentDPIUThreads – The concurrency limit for the DPIU processing queue.
- MaxConcurrentEventThreads – The concurrency limit for the DPAD event processing queue.

- `MaxConcurrentPublish` – The maximum number of concurrent Data Pool publication operations.
- `MaxConcurrentReadsPerTape` – The maximum number of concurrent tape read (stage) operations for a single tape.
- `MaxConcurrentRegister` - The maximum number of concurrent Data Pool registration operations.
- `MaxConcurrentValidate` – The maximum number of concurrent request validation operations.
- `MaxConsecutiveErrors` - The maximum number of consecutive errors or timeout conditions for a service before an alert will be raised.
- `MaxInsertRetries` - maximum number of times an insert should be tried again (-1 means forever).
- `MaxReadDrivesPerArchive` - maximum number of tape drives in use simultaneously.
- `MaxTapeMountPerRequest` - maximum number of tape mounts allowed per request.
- `NewActionCheckFrequency` – number of seconds before checking for new actions. DPAD always checks to determine whether we are out of actions that can be dispatched, so unless getting things queued up in memory is urgent, this could be a time interval of minutes.
- `NumOfAllowedCacheProcesses` - maximum number of insert processes that require access to cache (Obsolete in 7.20).
- `NumOfAllowedInsertProcesses` - maximum number of insert processes running at any time.
- `NumOfAllowedNonCacheProcesses` - maximum number of insert processes that require access to tape (Obsolete in 7.20).
- `OnTapeTimeLimit` - maximum time in hours operators are willing to wait for a DPIU process to complete when its files are not in cache. After the time limit, DPAD kills the process and retries the action.
- `OrderOnlyFSLabel` - order-only file system label.
- `RefreshRate` - DPM Home Page refresh rate in seconds.
- `RunAwayCheckFrequency` – number of seconds before checking again for runaway processes. It is recommended that `RunAwayCheckFrequency` not be much less than `InCacheTimeLimit`.
- `RunawayDuration` - maximum number of seconds to wait for an insert to complete before considering it a runaway (Obsolete in 7.20).

- **SizeOfInsertQueueList** - number of Data Pool Insert Queue entries that the DPM GUI can display on a page at any one time.
- **StartUpWait** - number of seconds to delay start-up while trying to clean out leftover DPIU processes (Obsolete in 7.20).

### 14.10.6.1 View DPM Configuration Parameter Values

---

- 1 Launch the **DPM GUI**
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Configuration Parameters** link.
  - The **List of Configuration Parameters** page is displayed (see Figure 14.10-13).
- 3 Observe data displayed on the **List of Configuration Parameters** page.
  - The table on the **List of Configuration Parameters** page has columns containing the following types of Data Pool configuration information:
    - **Parameter Name**
    - **Parameter Value** (including an entry field with current value, followed by a brief description of the parameter)
    - **Click on Box to Modify Parm** (containing a check box to mark the parameter for change)
  - The rows in the table indicate the current values and descriptions of the following types of parameters:
    - ActionQueueCleanupFrequency
    - AlertCleanupInterval
    - AlertNotifyEmailAddress
    - AlertRetentionPeriod
    - AlertRetryInterval
    - BatchSummaryAutoRefresh
    - ChecksumPercent
    - Clean703Orders
    - DPLRetentionPatchInstalled
    - DatabaseRetryCount
    - DatabaseRetryInterval
    - DefaultRetentionPeriod

- DefaultRetentionPriority
- DeleteCompletedActionsAfter
- DisplayAIPChunkSize
- FileSystemCheckInterval
- FileSystemRefreshRate
- FilterChecksumAIP
- FilterCopiedAIP
- FilterExtractedAIP
- FilterPendingAIP
- FilterValidAIP
- FreeSpaceResumePercent
- GranuleLockRetentionPeriod
- GranuleOMLockRetentionPeriod
- HEGCleanupAge
- InsertRetryWait
- MAX\_READ\_DRIVES\_x0xxgmn
- MaxConcurrentBandExtract
- MaxConcurrentDPIUThreads
- MaxConcurrentEventThreads
- MaxConcurrentPublish
- MaxConcurrentReadsPerTape
- MaxConcurrentRegister
- MaxConcurrentValidate
- MaxConsecutiveErrors
- MaxInsertRetries
- MaxReadDrivesPerArchive
- MaxTapeMountPerRequest
- NewActionCheckFrequency
- NumOfAllowedInsertProcesses

- OnTapeTimeLimit
  - OrdersOnlyFSLabel
  - RefreshRate
  - SizeOfInsertQueueList
  - There is an **Apply Change** button at the bottom of the page for implementing changes.
- 

Although most of the parameters managed on the **List of Configuration Parameters** page are not likely to be changed frequently, the operator may want to change some of them for tuning the Data Pool. Data Pool tuning parameters can be used to help meter the flow of data into the Data Pool and to adjust retention priority and duration to maintain optimum usage of Data Pool storage. To determine the best settings, it is necessary to monitor Data Pool inserts and disk space and adjust the parameters based on experience and projected functioning.

#### 14.10.6.2 Modify DPM Configuration Parameter Values

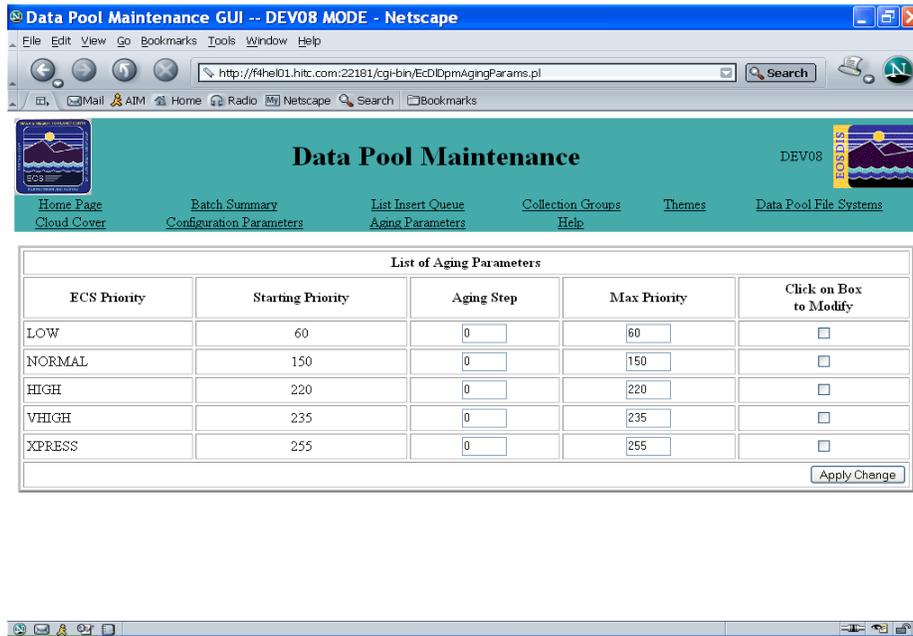
---

- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Configuration Parameters** link.
    - The **List of Configuration Parameters** page is displayed.
  - 3 If there is an option list for the parameter value to be changed, first click on the corresponding option button then click on the appropriate choice (e.g., **ON**)
    - Options are displayed (e.g., **ON** and **OFF**).
  - 4 If there is no option list for the parameter value to be changed, type the desired value in the corresponding text entry box.
  - 5 Click in the check box at the end of the row containing the parameter value to be modified.
    - The selected configuration information is marked for modification.
  - 6 Repeat Steps 3 through 5 for any additional parameter values to be modified.
  - 7 To implement the modification of parameter value(s) click on the **Apply Change** button.
    - The **List of Configuration Parameters** page is refreshed, the check box(es) is (are) unfilled, and the displayed **Parameter Value(s)** reflect(s) the change(s) implemented.
- 

#### 14.10.7 Aging Parameters

Figure 14.10-13 shows the **List of Aging Parameters** page, which is accessible from the **Aging Parameters** link on the **DPM GUI Home Page** (Figure 14.10-2). The page lists the starting

priority, aging step, and maximum priority associated with each ECS priority. The information is accessible to both full-capability and limited-capability operators.



**Figure 14.10-13. List of Aging Parameters Page**

### 14.10.7.1 View DPM Aging Parameter Values

- 1 Launch the **DPM GUI**.
  - For detailed instructions refer to the **Launch the DPM GUI** procedure (previous section of this lesson).
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Aging Parameters** link.
  - The **List of Aging Parameters** page is displayed (see Figure 14.10-13).
- 3 Observe data displayed on the **List of Aging Parameters** page.
  - The table on the **List of Aging Parameters** page has columns containing the following types of Data Pool configuration information:
    - **ECS Priority** (list of all ECS priorities)
    - **Starting Priority** (cannot be changed using the GUI)
    - **Aging Step** (includes an entry field with current value)
    - **Max Priority** (includes an entry field with current value)

- **Click on Box to Modify** parameter (containing a check box to mark the parameter for change)
  - The rows in the table indicate the current values and descriptions of the various ECS priorities, from **LOW** to **XPRESS**.
  - There is an **Apply Change** button at the bottom of the page for implementing changes.
- 

For the aging step, and maximum priority associated with each ECS priority there is a text box so the full-capability operator can assign a new value to the parameter when necessary. In addition, there is a check box that the full-capability operator uses to mark parameters with values to be modified. At the bottom of the page is an **Apply Change** button for implementing the change(s).

#### 14.10.7.2 Modify DPM Aging Parameter Values

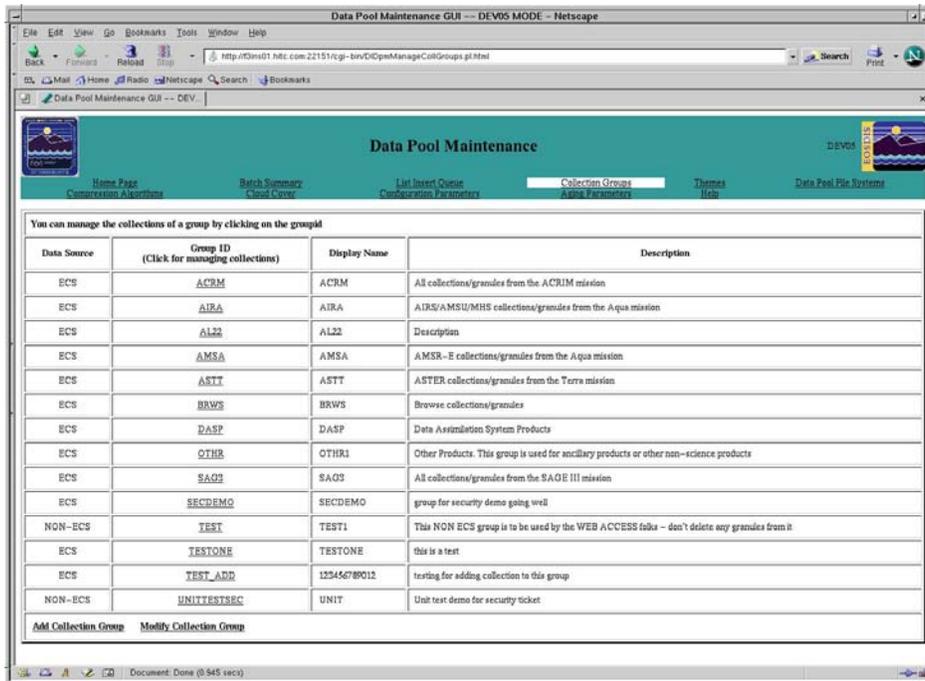
---

- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Aging Parameters** link.
    - The **List of Aging Parameters** page is displayed.
  - 3 To change the value associated with **Aging Step** and/or **Max Priority** for a particular ECS priority first type the desired value(s) in the corresponding text entry box(s).
  - 4 To continue the process of changing the value associated with **Aging Step** and/or **Max Priority** for a particular ECS priority click in the check box at the end of the row containing the parameter value(s) to be modified.
    - The selected configuration information is marked for modification.
  - 5 Repeat Steps 3 and 4 for any additional parameter values to be modified.
  - 6 To implement the modification of parameter value(s) click on the **Apply Change** button.
    - The **List of Aging Parameters** page is refreshed, the check box(es) is (are) unfilled, and the displayed **Aging Step** and **Max Priority** values reflect the change(s) implemented.
- 

#### 14.10.8 Collection Groups

Figure 14.10-14 illustrates the Collection Groups page and is accessible from the **Collection Groups** link on the **Home Page** (Figure 14.10-2). The page lists the collection groups, providing for each the **Data Source (ECS or Non-ECS)** (LaRC only), **Group ID**, **Display Name**, and a brief **Description** of the collection group.

Note: For the 7.21 LaRC delivery the Data Source field was added in order to accommodate Non-ECS data.

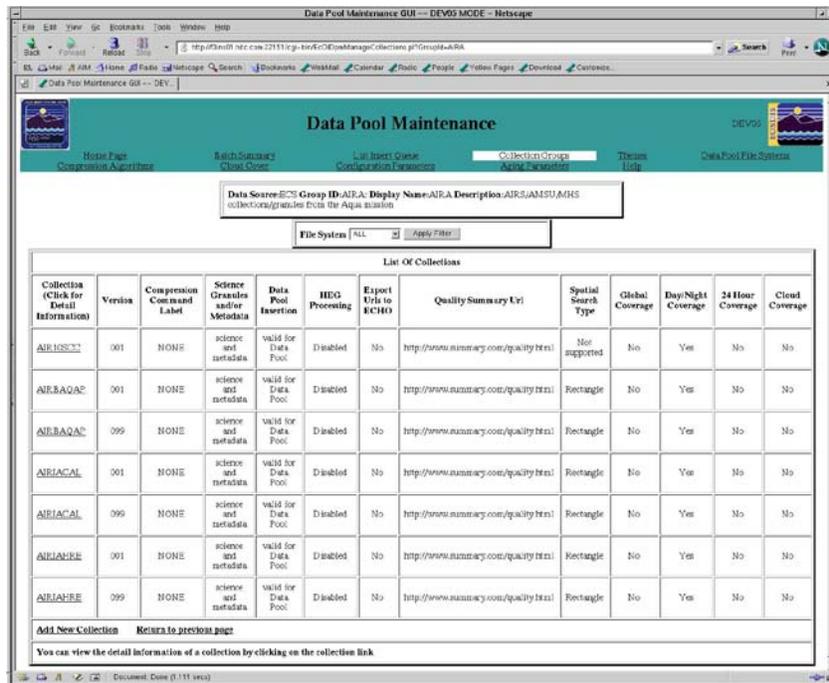


**Figure 14.10-14. Collection Groups Page**

The conceptual structure of the data pool is set up for each DAAC based on the collections and granules archived at the DAAC. Related collections are grouped in **Collection Groups** (e.g., ASTER collections and granules from the Terra mission, MODIS Oceans collections and granules from the Terra Mission, MISR collections and granules from the Terra mission, MODIS Snow and Ice collections and granules from the Terra mission). Each collection group initially consists of a number of collections that have been specified as valid for Data Pool insertion (i.e., granules of the data types in the collection may be inserted into the Data Pool).

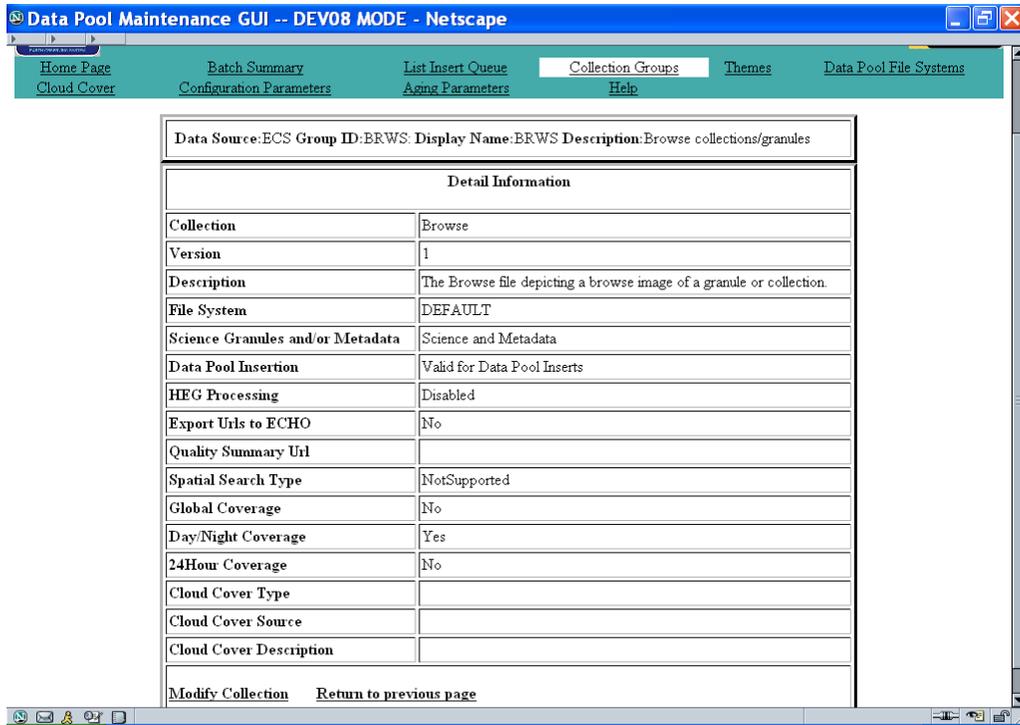
The Collection Groups page of the **DPM GUI** allows both full-capability operators and limited-capability operators to view collection groups. It also provides access to pages for viewing collections within a collection group. In addition, the page has links that allow a full-capability operator to modify or add a collection group or collection in the Data Pool database.

Figure 14.10-15 shows a **List of Collections** page obtained by clicking on one of the **Group ID** links on the Collection Groups page. The **List of Collections** page lists the collections in the collection group, providing for each collection information (as applicable) concerning the Version, Science Granules and/or Metadata, Data Pool Insertion, HDF-EOS to GeoTIFF Conversion Tool (HEG) Processing, Export URLs to ECHO, Quality Summary URL, Spatial Search Type, Global Coverage, Day/Night Coverage, 24-Hour Coverage, and Cloud Coverage characteristics of the collection.



**Figure 14.10-15. List of Collection Page**

At the bottom of the **List of Collections** page, there is a link that permits a full-capability operator to **Add New Collection** or **Delete a Collection**. A click on one of the **Collection (Click for Detail Information)** ID links brings up a collection **Detail Information** page (Figure 14.10-16) listing the same information for the collection as was displayed on the **List of Collection** page plus some additional information. The additional information includes a **Description**, **File System**, **Cloud Cover Type**, **Cloud Cover Source**, and **Cloud Cover Description**.



**Figure 14.10-16. Collection Detail Information Page**

The collection **List of Collections** and **Detail Information** pages provide a means of determining what collections within a collection group have been designated valid for Data Pool insertion and whether the insertion is for science granules and metadata or metadata only.

At the bottom of the **Detail Information** page, there is a link that permits a full-capability operator to modify a collection.

### 14.10.8.1 View Collection Groups

- 1 Launch the **DPM GUI**.
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Collection Groups** link.
  - The Collection Group page is displayed (see Figure 14.10-15).
- 3 Observe data displayed on the Collection Group page.
  - The table on the Collection Group page has columns containing the following types of collection group information:
    - **Data Source (LaRC only)** – Designates ECS or Non-ECS data.
    - **Group ID (Click for managing collections)**

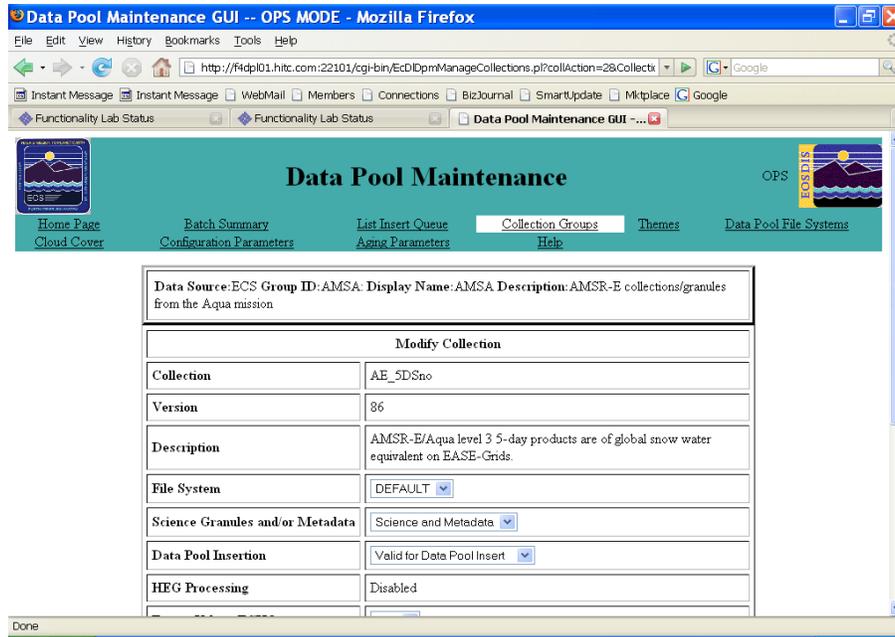
- **Display Name**
- **Description**
- The following links are available on the Collection Groups page:
  - **Group ID (Click for managing collections)** - Links to a **List of Collections** contained in that group
  - **Add Collection Group**
  - **Modify Collection Group**
- 4 To obtain more information about the collections in one of the groups, click on its link in the **Group ID (Click for managing collections)** column.
  - The **List of Collection** page is displayed (see Figure 14.10-15).
- 5 Observe data displayed on the **List of Collections** page.
  - Near the top of the **List of Collections** page contains the following basic collection group information:
    - **Data Source (LaRC only)** – Designates ECS or Non-ECS data.
    - **Group ID**
    - **Display Name**
    - **Description**
  - There is a file system filter (and associated **Apply Filter** button) for displaying data on the **Collection (Click for Detail Information)** column for all file systems or by individual file system.
  - The **List of Collection** page has columns containing the following types of collection group information:
    - **Collection (Click for Detail Information)** link
    - **Version**
    - **Science Granules and/or Metadata**
    - **Data Pool Insertion**
    - **HEG Processing**
    - **Export Urls to ECHO**
    - **Quality Summary Url**
    - **Spatial Search Type**
    - **Global Coverage**
    - **Day/Night Coverage**

- **24 Hour Coverage**
  - **Cloud Coverage**
  - **Nominal Coverage Rule**
  - **Check the Box to Delete Collection**
  - The following links are available on the **List of Collection** page:
    - Each collection listed in the **Collection** column links to a **Collection Detail** page.
    - **Add New Collection**
    - **Return to previous page**
- 6** To filter data displayed on the **List of Collections** page, click on the **File System** filter option button.
- Options are displayed.
- 7** Select a file system filter option click on the appropriate choice from the option list.
- 8** To implement the filtering of data displayed on the **Collection Detail** link, click on the **Apply Filter** button.
- The **Collection (Click for Detail Information)** column is displayed with the filtered collection group information.
- 9** To obtain more information about one of the collections in the collection group, click on its link in the **Collection (Click for Detail Information)** column.
- The **Detail Information** page (see Figure 14.10-16) for the selected collection is displayed.
- 10** Observe data displayed on the **Detail Information** page.
- Near the top of the **Detail Information** page is the following basic collection group information:
    - **Data Source**
    - **Group ID**
    - **Display Name**
    - **Description**
  - The **Detail Information** page has rows containing the following types of collection information:
    - **Collection**
    - **Version**
    - **Description**
    - **File System**

- **Science Granules and/or Metadata**
  - **Data Pool Insertion**
  - **HEG Processing**
  - **Export Urls to ECHO**
  - **Allow ordering and viewing of associated PH granule**
  - **Allow ordering and viewing of associated QA granule**
  - **Allow ordering of associated Browse granule**
  - **Quality Summary Url**
  - **Spatial Search Type**
  - **Global Coverage**
  - **Day/Night Coverage**
  - **24 Hour Coverage**
  - **Cloud Cover Type**
  - **Cloud Cover Source**
  - **Cloud Cover Description**
- The following links are available on the Collection **Information Detail** page:
    - **Modify Collection**
    - **Return to previous page**
- 11** To view a description for another collection in the same group first click on the **Return to previous page** link.
- The **List of Collections** page is displayed again.
- 12** To view a description for another collection in the same group return to Step 9.
- 13** To view a description for another collection in another group return to Step 2.
- 

The collection **List of Collections** and **Detail Information** pages provide a means of determining what collections within a collection group have been designated valid for Data Pool insertion and whether the insertion is for science granules and metadata or metadata only.

At the bottom of the **Detail Information** page, there is a link that permits a full-capability operator to modify a collection. Figure 14.10-17 shows the **Modify Collection** page obtained by clicking on a **Modify Collection** link. On this page, a full-capability operator can modify many of the characteristics of the collection then implement the changes with a click on the **Apply Change** button at the bottom.



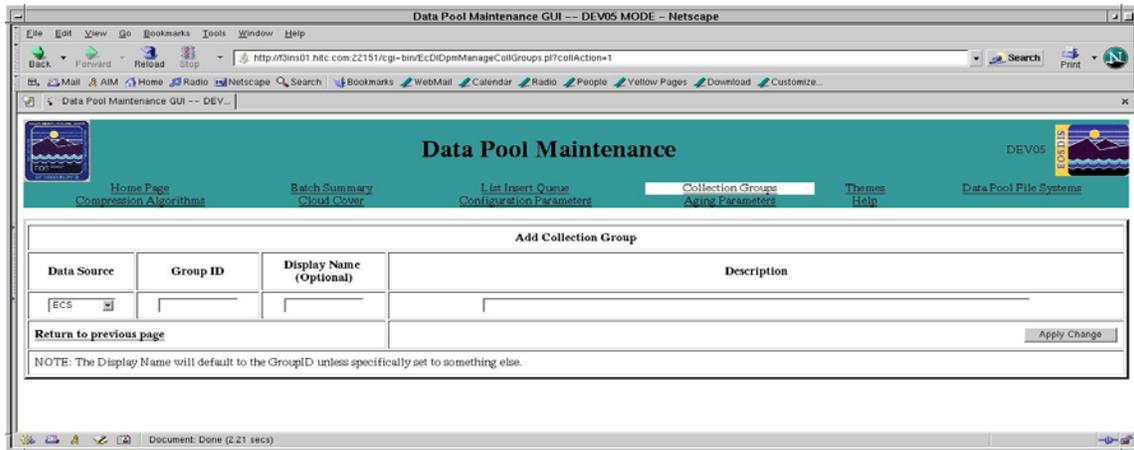
**Figure 14.10-17. Modify Collection Page**

From time to time, it may be necessary to add or modify a collection group (e.g., if a DAAC begins archiving data from a new instrument). If a collection group is to be added to the list of collection groups, it is necessary to use the **Add Collection Group** link at the bottom of the Collection Groups page. Full-capability operators (only) can use the procedure that follows to add a collection group (see Figure 14.10-18):

**NOTE:** Although the following procedure is applicable, most of the time new collection groups will be added only during releases of new software versions and you will not use this procedure often.

### **Caution**

The Add Collection Group function is to be exercised judiciously because the **DPM GUI** does not provide any means of deleting collection groups.



**Figure 14.10-18. Add Collection Group Page**

### 14.10.8.2 Modify Collection Groups

- 1 Launch the **DPM GUI**.
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Collection Groups** link.
  - The Collection Groups page is displayed.
- 3 Click on the **Modify Collection Group** link at the bottom of the page.
  - The **Modify Collection Group** page is displayed, providing a table of collection group information showing four columns **Group ID (Click to Manage Collections)**, **Display Name**, **Description**, and **Check box to Modify** (containing a check box to mark the collection group for change).
  - There is an **Apply Change** button at the bottom of the page for implementing changes.
- 4 To change the display name for the collection group, type the desired name in the **Display Name** field for the group ID.
  - The **Display Name** may have no more than 12 characters.
    - Valid characters include A-Z, 0-9, underscore and space.
- 5 To change the description of the collection group, type the desired description in the **Description** field for the group ID.
  - The **Description** may have no more than 255 characters.
- 6 Click in the check box at the end of the row containing collection group information to be modified.
  - The selected collection group information is marked for modification.
- 7 Repeat Steps 4 through 6 for any additional collection groups to be modified.

- 8 Click on the **Apply Change** button.
    - The revised collection group information is entered in the Data Pool database.
    - The Collection Group page is displayed with the modified collection group information.
- 

### 14.10.8.3 Add a Collection Group

---

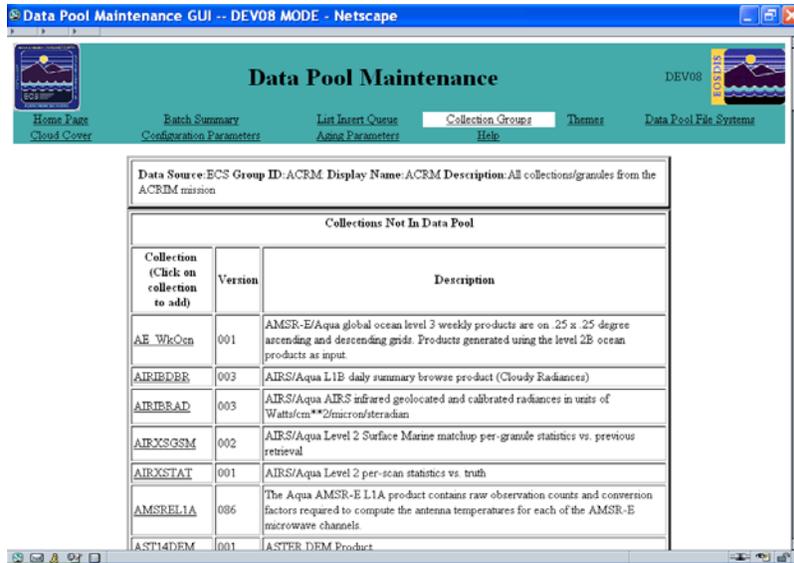
- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Collection Groups** link.
    - The Collection Groups page is displayed.
  - 3 Click on the **Add Collection Group** link at the bottom of the page.
    - The **Add Collection Group** page is displayed (see Figure 14.10-18) providing a page with three columns of text-entry fields, **Data Source**, **Group ID**, **Display Name**, and **Description**.
  - 4 Enter the **Data Source** (ECS or Non-ECS) for the new collection group in the **Data Source** field.
  - 5 Type a unique identifier for the new collection group in the **Group ID** field.
    - The **Group ID** may have no more than 12 characters.
      - Valid characters include A-Z, 0-9, and underscore.
    - The **Group ID** will be compared with the existing **Group IDs** to ensure that it is not a duplicate of another ID.
  - 6 To provide a display name that is different from the **Group ID** type a name in the **Display Name** field.
    - The **Display Name** is the name for the collection as displayed on the **Data Pool Web Access GUI**.
    - If no **Display Name** is entered, the **Group ID** will be used as the **Display Name**.
    - The **Display Name** may have no more than 12 characters.
      - Valid characters include A-Z, 0-9, underscore and space.
  - 7 Type the description for the new collection group in the **Description** field.
    - The **Description** may have no more than 255 characters.
  - 8 Click on the **Apply Change** button.
    - The new collection group information is entered in the Data Pool database.
    - The Collection Group page is displayed with the new collection group information.
-

#### 14.10.8.4 Delete a Collection

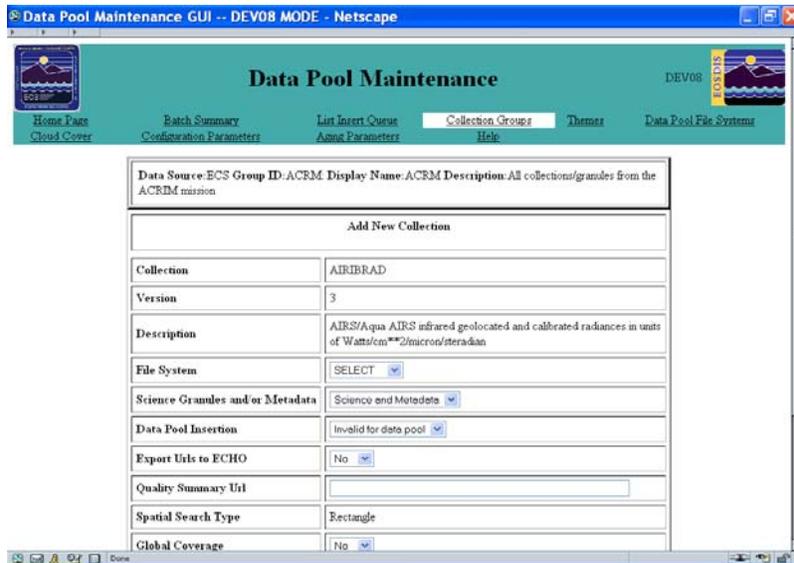
---

- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Collection Groups** link.
    - The Collection Groups Page is displayed
  - 3 Click on one of the Group ID links on the Collection Groups Page.
    - The List of Collections page is displayed.
  - 4 Scroll to the far right of the screen, and click on the **Delete Collection** box of the collection to be deleted.
    - A check mark is placed in the box..
  - 5 Click on the **Delete Collection** button (bottom of screen).
    - A confirmation window is displayed.
    - Select **OK** if you want to complete the delete process.
- 

Although an initial Data Pool structure is provided, not all collections are necessarily specified as eligible for Data Pool insertion. Based on experience, or on changes in demand, a DAAC may wish to add one or more collections to a data group. The procedure for adding ECS collections to a collection group is somewhat different from the procedure for adding a non-ECS collection to a collection group. If a full-capability operator clicks on an **Add New Collection** link at the bottom of a **List of Collections** page for an ECS collection, a **Collections Not in Data Pool** page (Figure 14.10-19) is displayed. The page lists ECS collections that are not currently part of a Data Pool collection group. The operator can select an ECS collection to add to the collection group by clicking on the link in the **Collection (Click on collection to add)** column of the table on the page. That causes an **Add New Collection** page (Figure 14.10-20) to be displayed. The Collection, Version, Description, and Spatial Search Type fields are filled in when the page comes up. The page has fields and option lists for entering the remaining data concerning the collection (e.g., File System, and Science Granules and/or Metadata). After the operator enters the appropriate data concerning the ECS collection, clicking on the **Apply Change** button at the bottom of the page applies the changes to the Data Pool database and refreshes the **List of Collections Group** page.



**Figure 14.10-19. Collections Not in Data Pool Page**



**Figure 14.10-20. Add New [ECS] Collection Page**

### 14.10.8.5 Add an ECS Collection to a Collection Group

---

- 1 Launch the **DPM GUI**.
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Collection Groups** link.
  - The Collection Group page is displayed.
- 3 Click on the **Group ID** link for the ECS collection group to which the collection is to be added.
  - The **List of Collections** page is displayed.
- 4 Click on the **Add New Collection** link at the bottom of the **List of Collections** page.
  - The **Collections Not in Data Pool** page is displayed (see Figure 14.9-19).
- 5 Click on the link in the **Collection (Click on collection to add)** column of the collection to be added to the collection group.
  - The **Add New Collection** page is displayed (see Figure 14.9-20).

**NOTE:** On the ECS collection version of the **Add New Collection** page the **Collection, Version, Description, and Spatial Search Type** fields are already filled in using information from the Data Pool database.

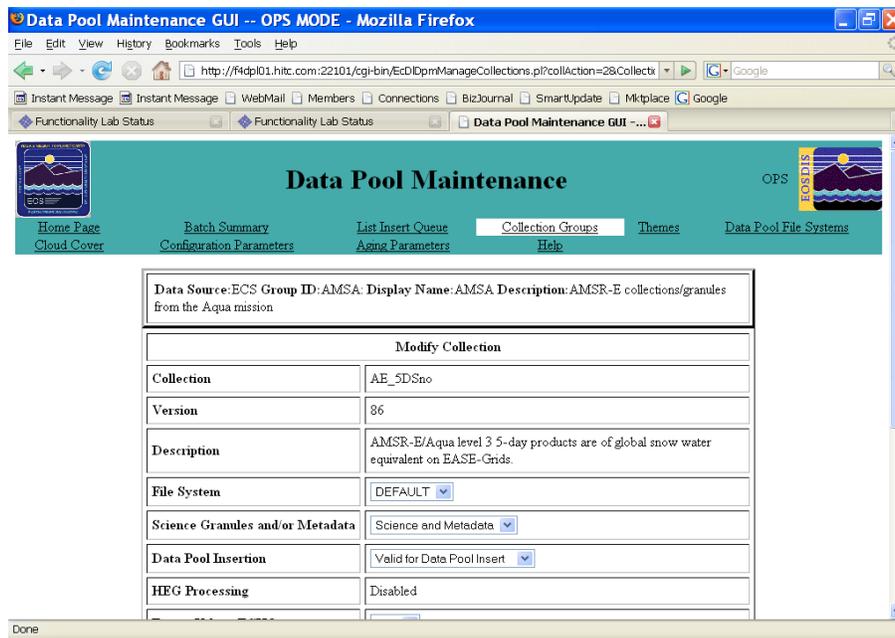
- 6 To select a file system option (if applicable), click on the appropriate choice from the **File System** option list.
- 7 To select a Science Granules and/or Metadata option, click on the appropriate choice from the Science Granules and/or Metadata option list.
  - **Science and Metadata** is the default option.
- 8 To select a data pool insertion option, click on the appropriate choice from the Data Pool Insertion option list.
  - **Invalid for data pool** is the default option.
  - **Valid for data pool** must be selected if the collection is to be eligible for insertion into the Data Pool.
- 9 To select an ECHO export option, click on the appropriate choice from the **Export Urls to ECHO** option list.
  - **No** is the default option.

**Yes** must be selected if collection URLs are to be eligible for export to ECHO.
- 10 To select an order/view PH granule, click on the appropriate choice from **Allow ordering and viewing of associated PH granule** option list.
  - **No** is the default option.
- 11 To select an order/view QA granule, click on the appropriate choice from **Allow ordering and viewing of associated QA granule** option list.
  - **No** is the default option.

- 12 To select an order browse granule, click on the appropriate choice from **Allow ordering of associated Browse granule** option list.
    - **No** is the default option.
  - 13 If the collection is to be linked to a quality summary web site, enter the URL in the **Quality Summary Url** text entry field.
    - Ensure that **http://** is included in the **Quality Summary Url** text entry field.
  - 14 To select a global coverage option, click on the appropriate choice from the **Global Coverage** option list.
    - **Yes** indicates no spatial searches for the collection.
    - **No** indicates that spatial searches are allowed for the collection.
  - 15 To select a day/night coverage option, click on the appropriate choice from the **Day/Night Coverage** option list.
    - **Yes** indicates that day/night searches are allowed for the collection.
    - **No** indicates that the collection is excluded from day/night searches.
  - 16 To select a 24-hour coverage option, click on the appropriate choice from the **24 Hour Coverage** option list.
    - **Yes** indicates that the collection is excluded from time of day searches.
    - **No** indicates that time of day searches are allowed for the collection.
  - 17 To select a cloud cover type and source option, click on the appropriate choice from the **Cloud Cover Type & Source** option list.
    - All cloud cover information in the Data Pool database is listed.
    - If the desired cloud cover type/source is not listed, it can be entered using the procedure **Add New Cloud Cover Information** (previous section of this lesson).
  - 18 To view details of cloud cover type and source, click on the **View Details** link adjacent to the **Cloud Cover Type & Source** option list.
  - 19 Click on the **Apply Change** button.
    - The new collection information is entered in the Data Pool database.
    - The **List of Collection** page is displayed with the new collection information.
- 

As part of managing the Data Pool storage and retention of data, making adjustments based on experience and/or changes in demand, it may be desirable to modify a collection. The modification may mean specifying that metadata only may continue to be inserted and science granules may no longer be inserted, or declaring the collection no longer valid for data pool insertion at all. Figure 14.10-21 shows the **Modify Collection** page that a full-capability operator obtains by clicking on the **Modify Collection** link at the bottom of the Detail Information page. On the **Modify Collection** page, the full-capability operator can modify the File System, Science Granules and/or Metadata, Data Pool Insertion, Quality Summary URL, DayNight Coverage, 24

Hour Coverage, Cloud Cover Type & Source. The operator implements the change(s) with a click on the **Apply Change** button at the bottom of the page.



**Figure 14.10-21. Modify Collection Page**

#### 14.10.8.6 Modify an ECS Collection

- 1 Launch the **DPM GUI**.
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Collection Groups** link.
  - The Collection Group page is displayed.
- 3 Click on the **Group ID** link for the ECS collection group to which the collection is to be added.
  - The **List of Collections** page is displayed.
- 4 Click on the desired link found in the **Collection (Click for Detail Information)** column.
  - The **Detail Information** page is displayed.
- 5 Click on the **Modify Collection** link.
  - The **Modify Collection** page is displayed (see Figure 14.10-21).

**NOTE:** On the ECS collection version of the **Modify Collection** page, the **Collection**, **Version**, **Description**, **Spatial Search Type**, **HEG Processing**, **Export Urls to ECHO** and **Global Coverage** fields cannot be edited.

- 6 To select a file system option (if applicable), click on the appropriate choice from the **File System** option list.
- 7 To select a Science Granules and/or Metadata option, click on the appropriate choice from the Science Granules and/or Metadata option list.
  - **Science and Metadata** is the default option.
- 8 To select a data pool insertion option, click on the appropriate choice from the Data Pool Insertion option list.
  - **Invalid for Data Pool** is the default option.

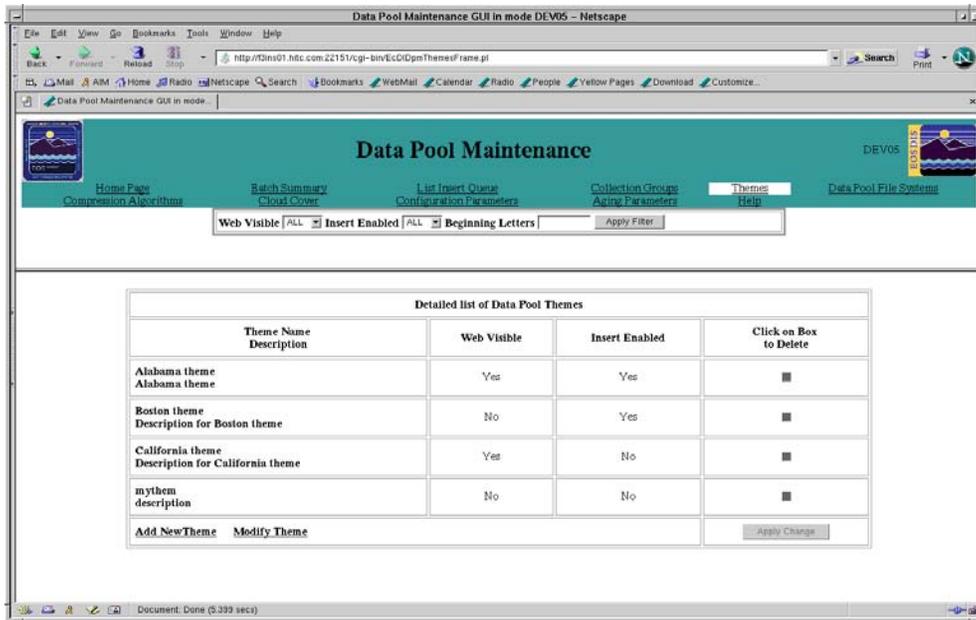
**Valid for Data Pool** must be selected if the collection is to be eligible for insertion into the Data Pool.
- 9 If the collection is to be linked to a quality summary web site, enter the URL in the **Quality Summary Url** text entry field.

Ensure that **http://** is included in the **Quality Summary Url** text entry field.
- 10 To select a day/night coverage option, click on the appropriate choice from the **Day/Night Coverage** option list.
  - **Yes** indicates that day/night searches are allowed for the collection.
  - **No** indicates that the collection is excluded from day/night searches.
- 11 To select a 24-hour coverage option, click on the appropriate choice from the **24 Hour Coverage** option list.
  - **Yes** indicates that the collection is excluded from time of day searches.
  - **No** indicates that time of day searches are allowed for the collection.
- 12 To select/update a cloud cover type and source option, click on the appropriate choice from the **Cloud Cover Type & Source** option list.
  - All cloud cover information in the Data Pool database is listed.
  - If cloud cover source needs updating, it can be chosen from the drop down list next to **Cloud Cover Type & Source**.
  - If the desired cloud cover type/source is not listed, it can be entered using the procedure **Add New Cloud Cover Information** (previous section of this lesson).
- 13 To view details of cloud cover type and source, click on the **View Details** link adjacent to the **Cloud Cover Type & Source** option list.
- 14 To select an order/view PH granule, click on the appropriate choice from **Allow ordering and viewing of associated PH granule** option list.
  - **Yes** indicates the collection(non-QA, non-PH, non-browse) is enabled for order/view PH
  - **No** indicates the collection(non-QA, non-PH, non-browse) is disabled for order/view PH

- 15 To select an order/view QA granule, click on the appropriate choice from **Allow ordering and viewing of associated QA granule** option list.
- **Yes** indicates the collection(non-QA, non-PH, non-browse) is enabled for order/view QA
  - **No** indicates the collection(non-QA, non-PH, non-browse) is disabled for order/view QA
- 16 To select an order browse granule, click on the appropriate choice from **Allow ordering of associated Browse granule** option list.
- **Yes** indicates the collection(non-QA, non-PH, non-browse) is enabled for order browse
  - **No** indicates the collection(non-QA, non-PH, non-browse) is disabled for order browse
- 14 Click on the **Apply Change** button.
- The new collection information is entered in the Data Pool database.
  - The **List of Collection** page is displayed with the new collection information.
- 

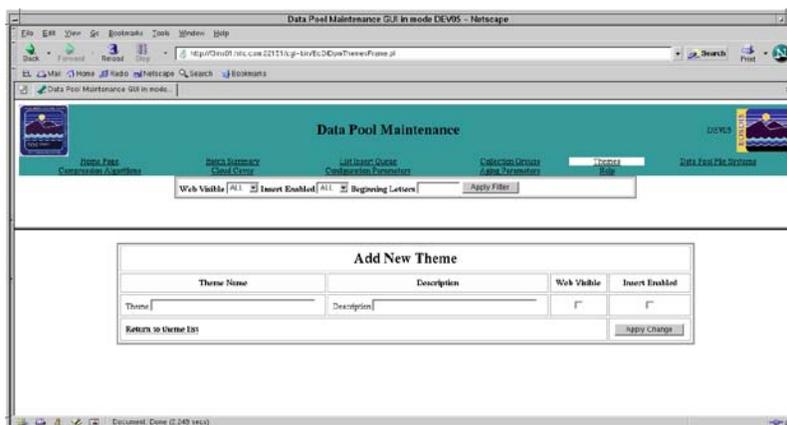
#### 14.10.9 Themes

Figure 14.10-22 illustrates the **Detailed List of Data Pool Themes** page. The page allows either the full-capability operator or the limited-capability operator to view a list of themes in alphabetical order. The list can be filtered using the option lists for **Web Visible** and **Insert Enabled**, and/or typing **Beginning Letters** (of the theme name). After selecting the options, a click on the **Apply Filter** button displays the filtered list of themes. The full-capability operator can delete a theme by selecting the corresponding check box and clicking on the **Apply Change** button. There are **Add New Theme** and **Modify Theme** links providing access to pages for managing those functions. After the operator completes adding a new theme or modifying a theme by clicking on the **Apply Change** button at the pages for those functions, the changes take effect in the Data Pool database and the changes are also reflected in the **Detailed List of Data Pool Themes** page.



**Figure 14.10-22. Detailed List of Data Pool Themes Page**

If a full-capability operator clicks on the **Add New Theme** link of the **Detailed List of Data Pool Themes** page shown in Figure 14.10-22, the **Add New Theme** page (Figure 14.10-23) is displayed. To specify a theme, the operator enters information in the fields provided for the purpose. **Theme Name** and **Description** are text entry fields. There are check boxes to specify whether the theme is valid for various options (i.e., **Web Visible** and **Insert Enabled**) or not. A click on the **Apply Change** button commits the changes to the Data Pool database and updates the **Detailed List of Data Pool Themes** page shown in Figure 14.10-22.



**Figure 14.10-23. Add New Theme Page**

Users may search the Data Pool for data associated with themes. As data are inserted into the Data Pool, it is possible to associate the data with themes. The **DPM GUI Detailed List of Data Pool Themes** page permits both full-capability and limited-capability operators users to view a list of Data Pool themes. In addition it has links that allow full-capability operators to add new themes, modify existing themes, or delete themes.

### 14.10.9.1 View a List of Themes

---

- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Themes** link.
    - The **Detailed List of Data Pool Themes** page is displayed (see Figure 14.10-23).
  - 3 Observe data displayed on the **Detailed List of Data Pool Themes** page.
    - The table on the Detailed List of Data Pool Themes page has columns containing the following types of Data Pool file system information:
      - **Theme Name/Description**
      - **Web Visible**
      - **Insert Enabled**
      - **Click on Box to Delete** (containing a check box to mark the theme for deletion)
    - There are theme filters (and associated **Apply Filter** button) for displaying data on the **Detailed List of Data Pool Themes** page depending on whether or not the theme.
      - Are web visible
      - Are insert enabled
      - Have certain letters at the beginning of the theme name.
    - Filters can be applied individually or in any combination.
    - The following links are available on the **Detailed List of Data Pool Themes** page:
      - **Add New Theme**
      - **Modify Theme**
    - There is an **Apply Change** button at the bottom of the page to implement the deletion of selected themes.
  - 4 To filter data displayed on the **Detailed List of Data Pool Themes** page use the **Filter a List of Themes** procedure (subsequent section of this lesson).
  - 5 If data displayed on the Detailed List of Data Pool Themes page were filtered, return to Step 3.
-

The procedure to **Filter a List of Themes** is subordinate to other theme-related procedures (i.e., **View a List of Themes**, **Modify a Theme**, and **Delete a Theme**.) Both full-capability and limited-capability operators users may filter data displayed on the Themes pages to which they have access.

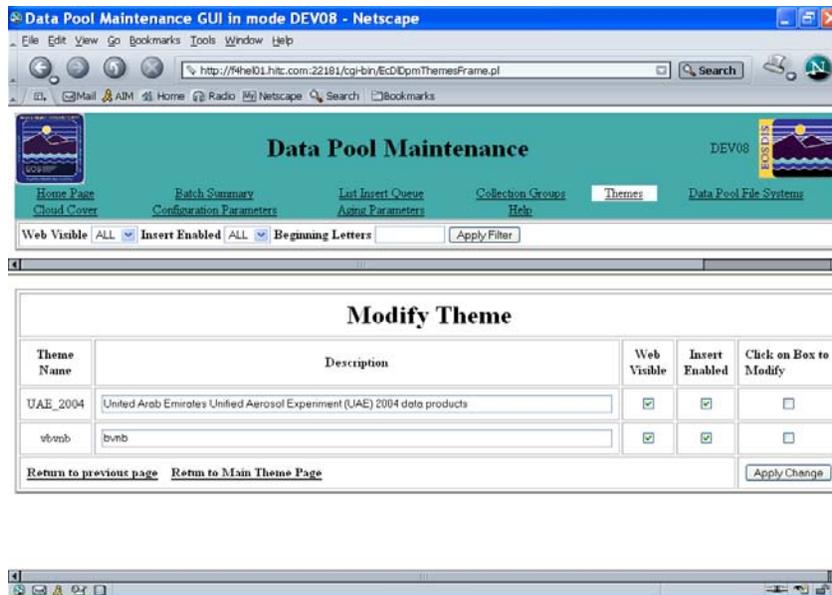
#### 14.10.9.2 Filter a List of Themes

---

- 1 To filter data displayed on one of the **Themes** pages on the basis of whether or not the themes are enabled for web drill-down, insertion in to the Data Pool, WCS accessibility, WMS accessibility, file format preconversion first click on one of the option buttons (in the filter area of the page):
    - **Web Visible**
    - **Insert Enabled**
  - 2 To continue the process of selecting a theme filter option click on the appropriate choice from the option list.
    - **Yes (View all themes enabled for the selected option)**
    - **No (View all themes disabled for the selected option)**
    - **ALL (View all themes regardless of whether the selected option is enabled or disabled)**
  - 3 Repeat Steps 1 and 2 as necessary to select additional filter options.
  - 4 To select a theme filter option on the basis of the beginning letters of the theme (if applicable) type the beginning letter(s) of the theme in the Beginning Letters text entry field.
  - 5 To implement the filtering of data displayed on one of the Themes pages click on the **Apply Filter** button.
    - The page is displayed with the filtered theme information.
  - 6 Return to the procedure that specified the **Filter a List of Themes** procedure.
- 

Full-capability operators can use the **DPM GUI** to modify a theme. This can be useful if, for example, it is noted that access frequency for granules referencing a theme has declined to the point that the thematic collection should be removed from the Data Pool, but there are a few web users that still use it. In that case, it may be appropriate to change the description of the theme to alert users that the theme will be phased out soon.

If a full-capability operator clicks on the **Modify Theme** link of the **Detailed List of Data Pool Themes** page shown in Figure 14.10-22, the **Modify Theme** page (Figure 14.10-24) is displayed. **Theme Name** is the only field that is not editable. The operator can modify the description of a theme by simply retyping in the text area. The operator also can change the various options (i.e., **Web Visible** and **Insert Enabled**) by selecting or deselecting the appropriate boxes. A click on the **Apply Change** button commits the changes to the Data Pool database and updates the **Detailed List of Data Pool Themes** page shown in Figure 14.10-22.



**Figure 14.10-24. Modify Theme Page**

### 14.10.9.3 Modify a Theme

- 1 Launch the **DPM GUI**.
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Themes** link.
  - The **Detailed List of Data Pool Themes** page is displayed.
- 3 Click on the **Modify Theme** link.
  - The **Modify Theme** page is displayed (see Figure 14.10-24).
- 4 To change the description of a theme (if applicable) type the desired description in the **Description** field for the theme name.
  - The **Description** may have no more than 255 characters.
- 5 To change the theme from enabled to disabled (or vice versa) for one of the options (i.e., **Web Visible, Insert Enabled**) (if applicable) click on the toggle button box in the corresponding column in the row for the theme.
  - A check mark in the box indicates that the theme is enabled for the corresponding option.
  - The absence of a check mark in the box indicates that the theme is not enabled for the corresponding option.

- 6 Click in the **Click on Box to Modify** check box at the end of the row containing the theme to be modified.
    - The selected theme is marked for modification.
  - 7 Repeat Steps 4 through 6 as necessary for any additional themes to be modified.
  - 8 To implement the modification of theme(s) click on the **Apply Change** button.
    - The theme information is entered in the Data Pool database.
    - The Detailed List of Data Pool Themes page is displayed with the modified theme information.
- 

Full-capability operators (only) can use the following procedure to add a theme:

#### **14.10.9.4 Add a Theme**

---

- 1 Launch the **DPM GUI**.
  - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Themes** link.
  - The **Detailed List of Data Pool Themes** page is displayed.
- 3 Click on the **Add New Theme** link.
  - The **Add New Theme** page is displayed (see Figure 14.10-24).
  - There is an **Apply Change** button at the bottom of the page for implementing changes.
  - The following link is available: **Return to theme list**.
- 4 Type a unique name for the theme in the **Theme Name** text entry field.
  - The Theme Name may have no more than 40 characters.
  - The Theme Name may not start with a number.
  - The Theme Name may not duplicate the name of a collection, an ESDT, or another theme.
- 5 To enter a description of the theme, type the desired description in the **Description** text entry field.
  - The **Description** may have no more than 255 characters.
- 6 To enable the theme for one of the options (i.e., **Web Visible, Insert Enabled**) (if applicable) click on the toggle button box in the corresponding area of the form.
  - A check mark in the box indicates that the theme is enabled for the corresponding option.
  - The absence of a check mark in the box indicates that the theme is not enabled for the corresponding option.
- 7 Repeat Step 6 as necessary to enable the theme for additional options.

- 8 Click on the **Apply Change** button.
    - The new theme information is entered in the Data Pool database.
    - The Detailed **List of Data Pool Themes** page is displayed with the new theme information.
- 

Full-capability operators (only) can use the following procedure to delete a theme:

#### 14.10.9.5 Delete a Theme

---

- 1 Launch the **DPM GUI**.
    - The **DPM GUI Home Page** is displayed.
  - 2 Click on the **Themes** link.
    - The **Detailed List of Data Pool Themes** page is displayed.
  - 3 Click in the **Click on Box to Delete** check box at the end of the row containing the theme to be deleted.
    - The selected theme is marked for deletion.
  - 4 Repeat Step 4 as necessary for any additional themes to be deleted.
  - 5 To implement the deletion of theme(s) click on the **Apply Change** button.
    - The theme deletion information is entered in the Data Pool database.
    - The **Detailed List of Data Pool Themes** page is displayed with the modified theme information.
- 

#### 14.10.10 Help

Figure 14.10-25 illustrates the **Help** page that allows both full-capability and limited-capability operators to obtain information on using the **DPM GUI**. The **Help** page describes the features of the other pages of the **DPM GUI**.

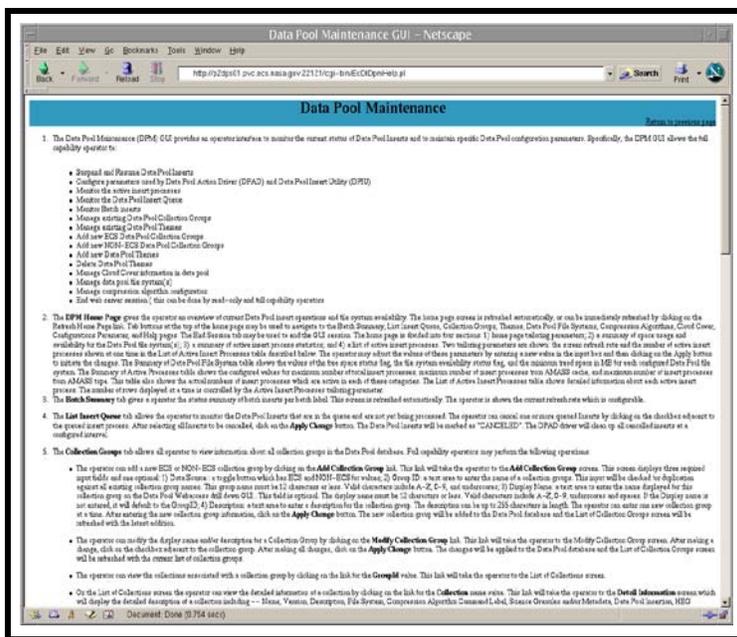


Figure 14.10-25. Help Page

## 14.11 Working with Data Pool Scripts

There are several Data Pool scripts that provide the operator with utilities or applications for managing Data Pool maintenance. These include:

- **Update Granule Utility:** a script to update granule expiration (extend the period of retention) and, optionally, retention priority, for selected science granules already in the Data Pool. The utility allows operators to extend the expiration of all granules associated with a particular thematic collection or upgrade their expiration priority.[With online archive capability, the utility is deprecated]
- **Data Pool Cleanup Orphan/Phantom Validation:** a script to check for orphans and phantoms in the Data Pool.
- **Data Pool SoftLink Check Utility:** a script to check for softlinks that do not point to a valid file.
- **Data Pool Cleanup Utility:** a script which removes granules in input file, regardless of whether they are public or hidden; will not remove granules in open orders.
- **Data Pool Online Archive Cleanup Utility:** a script to recover unprocessed granules that were left in tables DIBcpGransToDelete and DIBatchGransToDelete from the failure of previous running.
- **Data Pool Publish Utility:** a script to move granules to public Data Pool.

- Data Pool Unpublish Utility: a script to move granules from public Data Pool to hidden Data Pool.
- Data Pool Inventory Validation Utility: a script that compares Online Archive inventory (DPL db) with AIM inventory (AIM db); compares checksum in Online Archive (DPL db) with checksum in AIM inventory (AIM db).
- Data Pool Checksum Verification Utility (DPCV): a script that compares checksum in DPL database with checksum on disk in Online Archive. [Due to performance issue, this script is being replaced by Data Pool Checksum Verification Service (CVS)]
- Restore Online Archive from Tape Utility: a script to repair individual granules or files that are lost or damaged in the on-line archive but are still inventoried in the Data Pool database.
- Restore Tape from Online Archive Utility: a script to repair individual files that are lost or corrupted on tape based on the primary file instance that is present in the online archive. The files must be inventoried both in the AIM and DPL databases.
- Validation Tool: Archive Checksum Verification Utility (ACVU): a script that can identify corrupt files in the tape Archive. The utility uses a copy of the file in the Online Archive for validation.
- XML Check Utility (Xcu): a script to periodically check for corruption in the XML archive. In order to detect corruption, this utility verifies that the contents of the file are well-formed XML using xmllint.
- Data Pool Access Statistics Utility (DPASU): scripts for processing available Data Pool access logs to extract and summarize statistics on FTP and web access to data in the Data Pool. The statistics are stored in the Data Pool database to be used for producing tabular reports that can be loaded into a spreadsheet program for sorting, graphing, or other manipulation.
- Batch Insert Utility: a script for inserting non-ECS data and ECS data that are already in the archive into the Data Pool.[This script is being replaced by Publish Utility]
- Most Recent Data Pool Inserts Utility: normally runs as a cron job that lists the most recent additions to the Data Pool. If necessary, the utility can be run from the command line.
- Data Pool Collection-to-Group Remapping Utility: a command-line utility interface that is used for reassigning a Data Pool collection to a collection group other than the one to which it was originally assigned.
- Data Pool Move Collections Utility: a command-line interface to move collections from one file system to another. The file system move is implemented as a copy operation to the new collection directory, followed by removal of the old collection directory and its contents.
- Data Pool Hidden Scrambler Utility: a command-line utility for making the transition to or renaming (with encrypted names) hidden directories for order-only granules in the Data Pool.

- Data Pool Band Backfill Utility: a command-line tool that can correct band extraction problems that occurred during DPL registrations.
- Data Pool Remove Collection Utility: a command-line that provides a mechanism by which ECS Operations staff can remove collections from the Data Pool database that are no longer of interest to the end users.
- Data Pool Cloud Cover Utility: a command-line that enables the DAAC to correct the cloud cover (CC) information stored in the DPL database for public science granules, based on the cloud cover source that is associated to the desired collection via the DPL Maintenance GUI.
- DPL XML Check Utility (EcDIXcu.pl): a script to periodically check for corruption of the XML files in datapool. In order to detect corruption, this utility verifies that the contents of the file are well-formed XML using xmllint.

Table 14.11-1 provides an Activity Checklist for Data Pool Scripts addressed in this section.

**Table 14.11-1. Data Pool Scripts - Activity Checklist (1 of 2)**

| Order | Role               | Task                                                                                       | Section       | Complete? |
|-------|--------------------|--------------------------------------------------------------------------------------------|---------------|-----------|
| 1     | Archive Technician | Extend the Retention for Selected Science Granules Using the Update Granule Utility        | (P) 14.11.1.1 |           |
| 2     | Archive Technician | Running the Data Pool Cleanup Utility Manually                                             | (P) 14.11.2.1 |           |
| 3     | Archive Technician | Specify Data Pool Access Statistics Rollup Start Time and DPASU Execution with <i>cron</i> | (P) 14.11.3.1 |           |
| 4     | Archive Technician | Specify Data Pool Access Statistics Utility Execution from the Command Line                | (P) 14.11.3.2 |           |
| 5     | Archive Technician | Archive Access Statistics Using the Data Pool Archive Access Statistics Data Utility       | (P) 14.11.3.3 |           |
| 6     | Archive Technician | Delete Access Statistics Using the Data Pool Archive Access Statistics Data Utility        | (P) 14.11.3.4 |           |
| 7     | Archive Technician | Restore Access Statistics Using the Data Pool Archive Access Statistics Data Utility       | (P) 14.11.3.5 |           |
| 8     | Archive Technician | Running Batch Insert Utility                                                               | (P) 14.11.4.1 |           |
| 9     | Archive Technician | Running the Most Recent Data Pool Inserts Utility                                          | (P) 14.11.5.1 |           |
| 10    | Archive Technician | Running the Data Pool Collection-to-Group Remapping Utility                                | (P) 14.11.6.1 |           |
| 11    | Archive Technician | Running the Data Pool Move Collections Utility                                             | (P) 14.11.7.1 |           |

**Table 14.11-1. Data Pool Scripts - Activity Checklist (2 of 2)**

| Order | Role               | Task                                                          | Section        | Complete? |
|-------|--------------------|---------------------------------------------------------------|----------------|-----------|
| 13    | Archive Technician | Running the Data Pool Hidden Scrambler Utility in Rename Mode | (P) 14.11.8.1  |           |
| 14    | Archive Technician | Running the Data Pool Cleanup Orphan/Phantom Validation       | (P) 14.11.9.1  |           |
| 15    | Archive Technician | Running the Data Pool SoftLink Check Utility                  | (P) 14.11.10.1 |           |
| 16    | Archive Technician | Running the Data Pool Online Archive Cleanup Utility          | (P) 14.11.11.1 |           |
| 17    | Archive Technician | Running the Data Pool Publish Utility                         | (P) 14.11.12.1 |           |
| 18    | Archive Technician | Running the Data Pool Unpublish Utility                       | (P) 14.11.13.1 |           |
| 19    | Archive Technician | Running the Data Pool Inventory Validation Utility            | (P) 14.11.14.1 |           |
| 20    | Archive Technician | Running the Data Pool Checksum Verification Utility           | (P) 14.11.15.1 |           |
| 21    | Archive Technician | Running the Restore Online Archive from Tape Utility          | (P) 14.11.16.1 |           |
| 22    | Archive Technician | Running the Restore Tape from Online Archive Utility          | (P) 14.11.17.1 |           |
| 23    | Archive Technician | Running the Archive Checksum Verification Utility             | (P) 14.11.18.1 |           |
| 24    | Archive Technician | Running the XML Check Utility                                 | (P) 14.11.19.1 |           |
| 25    | Archive Technician | Running the Data Pool Band Backfill Utility                   | (P) 14.11.20.1 |           |
| 26    | Archive Technician | Running the Data Pool Remove Collection Utility               | (P) 14.11.21.1 |           |
| 27    | Archive Technician | Running the Data Pool Cloud Cover Utility                     | (P) 14.11.22.1 |           |
| 28    | Archive Technician | Running the Data Pool Checksum Verification Service Client    | (P) 14.11.23.1 |           |
| 29    | Archive Technician | Running the DPL XML Check Utility                             | (P) 14.11.24.1 |           |

### 14.11.1 Extending the Period of Retention for Granules in the Data Pool

Data Pool maintenance personnel can run the Update Granule Utility to update the expiration date for selected non-ECS science granules. This utility also permits modifying a granule's retention priority, which can affect how soon the Data Pool Cleanup Utility removes the granule from the Data Pool.

The Update Granule Utility permits updating granule information using a command-line interface. The following options may be used:

**-noprompt:** suppressing prompts and detailed information display.

**-theme:** specifies a valid theme name (i.e., a character string that matches an existing theme name in the Data Pool inventory).

A single non-ECS granule may be updated using manual input. Multiple non-ECS granule updates can be handled using an input file containing a list of granules to be updated, or by specifying a theme. The input file must be structured as a list of granules to be processed, one per line. Each line contains a granule ID (reflecting the Sybase entry in the Data Pool database), an expiration date, and (optionally) a new retention priority, the value of which may be null (i.e., left blank). The fields are separated by a single space. There should be no blank lines before the first or after the last granule in the list. The file contents should be similar to the following example.

```
GRANULE_ID_4832 EXP_DATE=2002/2/28 RETENTION=255
GRANULE_ID_4876 EXP_DATE=2002/2/28 RETENTION=200
GRANULE_ID_4883 EXP_DATE=2002/2/28 RETENTION=
GRANULE_ID_4937 EXP_DATE=2002/2/28
GRANULE_ID_4966 EXP_DATE=2002/2/28 RETENTION=255
```

When updating the granules associated with a theme, the utility updates the expiration date of a granule associated with that theme if and only if the new expiration date specified is later than the current expiration date of the granule. It updates the retention priority of a granule associated with that theme if and only if the new expiration priority specified is higher than the current retention priority of the granule.

The Update Granule Utility connects to the Data Pool database and calls Sybase stored procedures to perform the requested updates. Therefore, the utility runs only if the Data Pool database server is running and if the database is available. It also assumes the stored procedures are present. The Granule Update Utility may be run as a background process, with suppression of all warning/error messages and confirmation prompts if desired. When the utility is run, it writes information, any warnings, any errors, and messages to a log file about granules as they are updated.

Assume that a user contacts the DAAC with a request to update (extend) the expiration date to the end of February 2002 for selected granules in the Data Pool, and provides a list of granule IDs for the selected granules. The following procedure is applicable.

#### **14.11.1.1 Extend the Retention for Selected Science Granules Using the Update Granule Utility**

---

- 1** Log in at the machine on which the Update Granule Utility is installed (e.g., x4dpl01).
- 2** To change to the directory for starting the Update Granule Utility, type:  
**cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.
- 3** At the UNIX prompt, type the command to start the Update Granule Utility, in the form **EcDIUpdateGranule.pl <command line parameters>** then press the **Return/Enter** key.
  - For this exercise, use the following command:

**EcDIUpdateGranule.pl <MODE> -file <tr\_list>**

The first command-line parameter specified must be <MODE>, a valid, existing Data Pool mode (e.g., OPS, TS1, TS2).

- The following six permutations are valid command-line entries for initiating the Update Granule utility:
  - **EcDIUpdateGranule.pl <MODE> -file <filename>** (to update non-ECS granules listed in an input file named <filename> while displaying all summary information to the operator, and asking confirmation of the update).
  - **Ec DIUpdateGranule.pl <MODE> -grnid <granuleID> -exp <expiration date> [-ret <retention priority>]** (to update a non-ECS granule identified by its <granuleID> with a new expiration date and, optionally, a new retention priority while displaying all summary information to the operator, and asking confirmation of the update).
  - **EcDIUpdateGranule.pl <MODE> -noprompt -file <filename>** (to update non-ECS granules listed in an input file named <filename> with no confirmation or information displayed to the operator).
  - **EcDIUpdateGranule.pl <MODE> -noprompt -grnid <granuleID> -exp <expiration date> [-ret <retention priority>]** (to update a non-ECS granule identified by its <granuleID> with a new expiration date and, optionally, a new retention priority with no confirmation or information displayed to the operator).
  - **EcDIUpdateGranule.pl <MODE> -theme <themename> -exp <expiration date> [-ret <retention priority>]** (to update a non-ECS granule identified by its <themename> with a new expiration date and, optionally, a new retention priority while displaying all summary information to the operator, and asking confirmation of the update).
  - **EcDIUpdateGranule.pl <MODE> -noprompt -theme <themename> -exp <expiration date> [-ret <retention priority>]** (to update a non-ECS granule identified by its <themename> with a new expiration date and, optionally, a new retention priority with no confirmation or information displayed to the operator).

- The utility executes and displays a confirmation prompt similar to the following:

**You are about to start updating granules.**

-----

**Total number of granules: 11**

**Total size of granules: 8.61339673772454 MB**

**Do you wish to continue processing the update? [y/n]y**

4 Type y and then press the **Return/Enter** key.

- The utility completes execution and displays output similar to the following:

**Update completed.**

**Please check the database to ensure proper completion.**

## Update took 2 seconds to complete

### Gracefully exiting...

- To check the database, have the Database Administrator use isql commands on the Inventory database host to query the DIGranuleExpirationPriority table. It may also be useful to examine the Update Granule Utility log file to determine whether there were any problems with the execution. To examine that log file, go to Steps 5 and 6.
- 5** To change to the directory containing the Update Granule Utility log file and other log files, type the following:  
**cd /usr/ecs/<MODE>/CUSTOM/logs** and then press the **Return/Enter** key.
- The working directory is changed to /usr/ecs/<MODE>/CUSTOM/logs.
- 6** To examine the Update Granule Utility log file, type **pg EcDIUpdateGranule.log** and then press the **Return/Enter** key.
- The first page of the log file is displayed; additional sequential pages can be displayed by pressing the **Return/Enter** key at the “:” prompt. It is also possible to search forward by typing /<search item>. For example, to search the log file for reference to one of the granules updated, type /<granuleID> and then press the **Return/Enter** key.
  - Although this procedure is written for the **pg** command, any UNIX editor or visualizing command (e.g., **vi**, **view**, **more**, or **tail**) can be used to review the log.
  - The log entries have a time and date stamp; about the time that the update was executed, the log should show entries similar to the following:  
**2001/11/29 15:52:50.814:Update started...**  
**2001/11/29 15:52:50.964:Granule 4871 updated**  
**2001/11/29 15:52:51.083:Granule 4954 updated**  
**2001/11/29 15:52:51.212:Granule 4955 updated**  
**2001/11/29 15:52:51.346:Granule 4956 updated**  
**2001/11/29 15:52:51.409:Granule 4957 updated**  
**2001/11/29 15:52:51.688:Granule 4959 updated**  
**2001/11/29 15:52:51.778:Granule 4961 updated**  
**2001/11/29 15:52:51.998:Granule 4963 updated**  
**2001/11/29 15:52:52.107:Granule 4963 updated**  
**2001/11/29 15:52:52.394:Granule 4964 updated**  
**2001/11/29 15:52:52.569:Granule 4966 updated**  
**2001/11/29 15:52:52.590:Update ended.**  
**2001/11/29 15:52:52.608:This update took approximately 2 seconds**
  - If the log indicates errors or warnings, it may be necessary to correct the condition identified in the entry (e.g., edit the data in the granule list in the input file) and run the utility again. Specific error entries depend on the error that occurred; examples of error entries in the log may be similar to the following:  
**4959 AST\_04 1 0.03962299 Jul 30 2001 12:00AM Feb 2 1998 11:59PM**  
**255 2**

**Warning: The new expiration date for the above granule is less than or equal to today's date.**

**DATABASE ERROR:Server message number=120001 severity=16 state=1**

**line=33 server= x4oml01\_srvr procedure=ProcSelectGrExpiration**

**text=ProcSelectGrExpiration: Requested granule id not in database.**

**2001/11/29 15:50:36.647:Sybase Lookup ==> ERRORS WERE FOUND WITH GRANULE "4654". (It may not exist or contains the wrong format).**

**2001/11/29 15:50:36.663:**

**EcDIUpdateGranule\_1.pl aborted due to insufficient processing data: All the granule triplets had errors.**

---

## **14.11.2 Running the Data Pool Cleanup Utility (DELETED)**

---

### **14.11.3 Running the Data Pool Access Statistics Utility**

The Data Pool Access Statistics Utility (DPASU) parses logs of the Data Pool Web Access service and the FTP access service and stores the results in tables in the Data Pool database. The DPASU is a command-line utility that permits an option of entering input parameters. It is intended to be run with *cron* to cover an arbitrary 24-hour period starting at a time specified as a configuration parameter in a configuration file. However, an operator may run the utility from the command line specifying a start date as an input parameter to cover a period other than the normal 24-hour period addressed by *cron* or to cover that normal period if *cron* failed to process the logs for that period.

There are two versions of the DPASU, one for each type of log processed. The script named **EcDIRollupWebLogs.pl** runs on the Data Pool Web Access server and processes its log; its configuration file is **EcDIRollupWebLogs.CFG**. The script named **EcDIRollupFwFtpLogs.pl** runs on a server with access to SYSLOG with FTP access entries; its configuration file is **EcDIRollupFwFtpLogs.CFG**. These scripts capture data on downloads from the Data Pool, including date and time of access, path and file name of the file, and size of the file. The captured data are written to a temporary "flat file" -- a tab-delimited text file -- stored in the directory `<ECS_HOME>/<MODE>/CUSTOM/data/DPL/`. The flat file is then exported to Sybase and stored in a table. The DPASU calls Sybase stored procedures to generate a separate rollup table, removes the flat file, and enters a record in a separate table identifying which periods have been rolled up in order to prevent inadvertent reprocessing of that period.

To prevent potential table locking, *cron* runs of the DPASU scripts should be separated so that they are not both running concurrently (e.g., separate their start times by at least 20 minutes). Use the following procedure to specify a 1:00 a.m. start time for the rollup and add a line to the *crontab* files to run the DPASU for the OPS mode beginning at 2:00 a.m. every day with a 20-minute separation between the scripts.

### 14.11.3.1 Specify Data Pool Access Statistics Rollup Start Time and DPASU Execution with *cron*

---

- 1 Log in at the host for EcDIRollupWebLogs.pl and its configuration file (e.g., x4dpl01).
- 2 To change to the directory containing the configuration file, type the following command:  
**cd /usr/ecs/OPS/CUSTOM/cfg** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/OPS/CUSTOM/cfg**.
- 3 To look at the Rollup Start Time specified in the configuration file, type **vi EcDIWebRollup.CFG** and then press the **Return/Enter** key.
  - The contents of the file are displayed, and the last line of the file indicates the start time in format similar to the following:  
**ROLLUP\_START\_TIME=3:00**
  - If the start time is correct, exit **vi** by typing **:q!** and pressing the **Return/Enter** key; then go to Step 10. Otherwise, to change the time, execute Steps 4 - 9.
- 4 Use the arrow keys on the keyboard to move the cursor down to the line specifying the **ROLLUP\_START\_TIME** and to move it to the right until it is located over the first character in the time value.
  - The cursor is moved to the start time location; the line should look similar to the following:  
**ROLLUP\_START\_TIME=3:00**
- 5 Type **x** to delete the number under the cursor.
  - The number is deleted; the line should look similar to the following.  
**ROLLUP\_START\_TIME= 00**
- 6 Type **i** to put the **vi** editor into the insert mode.
  - The **vi** editor is in the insert mode, but no feedback is provided.
- 7 Type **1**.
  - The typed entry appears to the left of the cursor.
- 8 Press the **Esc** key.
  - The cursor moves one character to the left and the **vi** editor is in the command mode.
- 9 Type **ZZ** (be sure to use upper case).
  - The file is saved and the UNIX prompt is displayed.
- 10 To ensure that the **crontab** command launches the **vi** editor, type **setenv EDITOR vi** and then press the **Return/Enter** key.
  - It may be desirable to include this command in the operator's **.cshrc** file to set the **crontab** editor to **vi** as part of the environmental settings normally used routinely.

- 11 Type **crontab -e** and then press the **Return/Enter** key.
  - The contents of the file are displayed, and the cursor is displayed on the first character at the upper left corner of the file.
  - If the operator has no **crontab** file on the current platform, this command opens a new one for editing.
- 12 If necessary, use the down arrow key on the keyboard to move the cursor down to a blank line.
  - The cursor is displayed at the beginning of the selected line.
- 13 Type **i** to put the **vi** editor into the insert mode.
  - The **vi** editor is in the insert mode, but no feedback is provided.
- 14 Type **0 2 \* \* \* /usr/ecs/OPS/CUSTOM/utilities/EcDIRollupWebLogs.pl OPS -noprompt**.
- 15 Press the **Esc** key.
  - The cursor moves one character to the left and the **vi** editor is in the command mode.
- 16 Type **:wq** and then press the **Return/Enter** key.
  - UNIX displays a message identifying the number of lines and characters in the **crontab** file (stored in the directory **/var/spool/cron/crontabs**) and then displays the UNIX prompt.
- 17 Log in at the host for EcDIRollupFtpLogs.pl and its configuration file (e.g., x4eil01).
- 18 To change to the directory containing the configuration file, type the following command: **cd /usr/ecs/OPS/CUSTOM/cfg** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/OPS/CUSTOM/cfg**.
- 19 To look at the Rollup Start Time specified in the configuration file, type the following: **vi EcDIRollupWuFtpLogs.CFG** and then press the **Return/Enter** key.
  - The contents of the file are displayed, and the last line of the file indicates the start time in format similar to the following:  
**ROLLUP\_START\_TIME=3:00**
  - If the start time is correct, exit **vi** by typing **:q!** and pressing the **Return/Enter** key; then go to Step 21. Otherwise, to change the time, execute Step 20.
- 20 Repeat Steps 4-9 to change the time in **EcDIRollupWuFtpLogs.CFG**.
- 21 To ensure that the **crontab** command launches the **vi** editor, type **setenv EDITOR vi** and then press the **Return/Enter** key.
  - It may be desirable to include this command in the operator's **.cshrc** file to set the **crontab** editor to **vi** as part of the environmental settings normally used routinely.
- 22 Type **crontab -e** and then press the **Return/Enter** key.
  - The contents of the file are displayed, and the cursor is displayed on the first character at the upper left corner of the file.

- If the operator has no **crontab** file on the current platform, this command opens a new one for editing.
- 23** If necessary, use the down arrow key on the keyboard to move the cursor down to a blank line.
- The cursor is displayed at the beginning of the selected line.
- 24** Type **i** to put the **vi** editor into the insert mode.
- The **vi** editor is in the insert mode, but no feedback is provided.
- 25** Type **20 2 \* \* \* /usr/ecs/OPS/CUSTOM/utilities/EcDIRollupWuFtpLogs.pl OPS -noprompt**.
- The typed entry appears to the left of the cursor.
- 26** Press the **Esc** key.
- The cursor moves one character to the left and the **vi** editor is in the command mode.
- 27** Type **:wq** and then press the **Return/Enter** key.
- UNIX displays a message identifying the number of lines and characters in the **crontab** file (stored in the directory **/var/spool/cron/crontabs**) and then displays the UNIX prompt.
- 

Although the Data Pool Access Statistics Utility scripts are intended to be run with **cron**, if it is necessary to run them from the command line, it is possible to do so. For example, if **cron** fails to complete successfully for any reason, no entry is made into the record table to indicate that a period was processed. In that event, the statistics can be captured for the missing interval by running the utility manually.

There are seven command-line parameters for use with the utility scripts (see 609 document information):

- The **<MODE>** parameter indicates the mode (must specify a valid directory path) in which the script is to run; it is mandatory, unlabeled, and must be the first parameter following the command.
- The **-noprompt** parameter optionally specifies suppression of output to the screen.
- The **-nodelete** parameter optionally prevents the flat file from being deleted upon completion of the run.
- The **-flatfile <path/file>** parameter optionally provides an alternative path/file name for the flat file produced by the parser (useful only with the **-nodelete** option).
- The **-ftp <path/file>** parameter optionally indicates an alternative ftp log path/file(s) to be used instead of the configured default path/file (for the **EcDIRollupWuFtpLogs.pl** script only). Wildcards may be used, but must be escaped (i.e., preceded with a \).

- The **-web <path/file>** parameter optionally indicates an alternative web log path/file(s) to be used instead of the configured default path/file (for the **EcDIRollupWebLogs.pl** script only). Wildcards may be used, but must be escaped (i.e., preceded with a `\`).
- The **-start <date>** parameter optionally indicates an alternative start date for the rollup period, using the format MM/DD, and may be used to process a previously uncovered period.

With the exception of the mandatory **<MODE>** parameter, which must appear first after the command, the other parameters may be used in various orders and combinations. For example, to run without screen prompts or information, starting from December 22, and to retain the flat file, the command for accumulating statistics on web access should be entered as follows:

**EcDIRollupWebLogs.pl OPS -noprompt -nodelete -start 12/22.**

To run with normal screen information display, starting from February 15, but using an alternative file with wildcards for the web log, the command should be similar to the following:

**EcDIRollupWebLogs.pl OPS - start 2/15 -web /usr/var/\\*.log.**

Use the following procedure to run the Data Pool Access Statistics Utility scripts from the command line, with normal screen information display.

### 14.11.3.2 Specify Data Pool Access Statistics Utility Execution from the Command Line

- 1 Log in at the host for EcDIRollupWebLogs.pl and its configuration file (e.g., x4eil01).
- 2 To change directory to the directory containing the script, type the command **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.
- 3 Type **EcDIRollupWebLogs.pl <MODE>** and then press the **Return/Enter** key.
  - The utility runs and displays information to the screen as it executes, in form similar to the following:

#### A Data Pool product

```

\	_V	/ _\			
0	/ 0 \| \				
		/ \ \| \ \|			
_/	_/	_/	_/	_/	

```

#### Data Pool Access Statistics Utility

Connecting to database...

The DPASU will examine the logs for access entries between the following times:

| Month     | Day | Hour | Minute |
|-----------|-----|------|--------|
| -----     |     |      |        |
| START: 11 | 26  | 03   | 00     |
| END: 11   | 27  | 02   | 59     |

-----

START: 11 26 03 00

END: 11 27 02 59

Checking for already covered rollup periods...



**Rollup successful!**  
**Removing flat file...OK**  
**Gracefully exiting...**

---

The three remaining utilities are shell scripts for archiving, deleting, and restoring information in database tables populated by the DPASU. The **Data Pool Archive Access Statistics Data Utility** is run from the command line as needed or desirable to connect to the Data Pool database and write granule access data for a specified time range from the DIGranuleAccess, DIGranuleSubscription, and DIAccessRollup tables to an ASCII file. Once this is done, the operator can run the **Data Pool Delete Access Statistics Data Utility** from the command line to delete the archived data from the Data Pool database. If it is desirable to restore deleted data to the database, the **Data Pool Restore Access Statistics Data Utility** can be run from the command line to restore the data. The following procedures are applicable.

### **14.11.3.3 Archive Access Statistics Using the Data Pool Archive Access Statistics Data Utility**

---

- 1** Log in at the host for the Data Pool database (e.g., x4db103).
- 2** To change directory to the directory containing the Data Pool Archive Access Statistics Data Utility, type the following:  
**cd /usr/ecs/<MODE>/CUSTOM/dbms/DPL** and then press the **Return/Enter** key.
  - The working directory is changed to **cd /usr/ecs/<MODE>/CUSTOM/dbms/DPL**.
- 3** Type the following:  
**DI DbArchiveAccessStat <MODE> <STARTDATE> <STOPDATE> <ARCHIVEDIR> <USERNAME> <SERVER> <DBNAME>** and then press the **Return/Enter** key.
  - **<MODE>** is the mode in which the utility is being executed (e.g., OPS, TS1, TS2).
  - **<STARTDATE>** is the start date time range, in format **yyyymmdd**, for the data to be archived.
  - **<STOPDATE>** is the stop date time range, in format **yyyymmdd**, for the data to be archived.
  - **<ARCHIVEDIR>** is the absolute path where the generated ASCII files are to be stored.
  - **<USERNAME>** is the Sybase login name.
  - **<SERVER>** is the Sybase Server for the Data Pool database (e.g., x4db103\_srvr).
  - **<DBNAME>** is the name of the Data Pool database (e.g., DataPool OPS).
  - The script displays a prompt for entry of the password for the Sybase login.

**NOTE:** The step that follows may require input from the Database Administrator.

- 4 Type *<password>* and then press the **Return/Enter** key.
    - The script runs and the Archive Access Statistics Utility log file **DIDbArchiveAccessStat.log** records errors, warnings, and information about utility events. The log is written to the directory **/usr/ecs/<MODE>/CUSTOM/logs**.
- 

To run the Data Pool Delete Access Statistics Data Utility, use the following procedure.

#### 14.11.3.4 Delete Access Statistics Using the Data Pool Delete Access Statistics Data Utility

---

- 1 Log in at the host for the Data Pool database (e.g., x4dbl03).
- 2 To change directory to the directory containing the Data Pool Delete Access Statistics Data Utility, type the following:  
**cd /usr/ecs/<MODE>/CUSTOM/dbms/DPL** and then press the **Return/Enter** key.
  - The working directory is changed to **cd /usr/ecs/<MODE>/CUSTOM/dbms/DPL**.
- 3 Type the following:  
**DIDbDeleteAccessStat <MODE> <STARTDATE> <STOPDATE> <USERNAME> <SERVER> <DBNAME>** and then press the **Return/Enter** key.
  - **<MODE>** is the mode in which the utility is being executed (e.g., OPS, TS1, TS2).
  - **<STARTDATE>** is the start date time range, in format **yyyymmdd**, for the data to be deleted.
  - **<STOPDATE>** is the stop date time range, in format **yyyymmdd**, for the data to be deleted.
  - **<USERNAME>** is the Sybase login name.
  - **<SERVER>** is the Sybase Server for the Data Pool database (e.g., x4dbl03\_srvr).
  - **<DBNAME>** is the name of the Data Pool database (e.g., DataPool OPS).
  - The script displays a prompt for entry of the password for the Sybase login.

**NOTE:** The step that follows may require input from the Database Administrator.

- 4 Type *<password>* and then press the **Return/Enter** key.
    - The script runs and the Delete Access Statistics Utility log file **DIDbDeleteAccessStat.log** records errors, warnings, and information about utility events. The log is written to the directory **/usr/ecs/<MODE>/CUSTOM/logs**.
-

To run the Data Pool Restore Access Statistics Data Utility, use the following procedure.

### 14.11.3.5 Restore Access Statistics Using the Data Pool Restore Access Statistics Data Utility

---

- 1 Log in at the host for the Data Pool database (e.g., x4dbl03).
- 2 To change directory to the directory containing the Data Pool Restore Access Statistics Data Utility, type `cd /usr/ecs/<MODE>/CUSTOM/dbms/DPL` and then press the **Return/Enter** key.
  - The working directory is changed to `cd /usr/ecs/<MODE>/CUSTOM/dbms/DPL`.
- 3 Type `DIDbRestoreAccessStat <MODE> <STARTDATE> <STOPDATE> <ARCHIVEDIR> <USERNAME> <SERVER> <DBNAME>` and then press the **Return/Enter** key.
  - `<MODE>` is the mode in which the utility is being executed (e.g., OPS, TS1, TS2).
  - `<STARTDATE>` is the start date time range, in format `yyyymmdd`, for the data to be restored.
  - `<STOPDATE>` is the stop date time range, in format `yyyymmdd`, for the data to be restored.
  - `<ARCHIVEDIR>` is the absolute path of the storage location for the ASCII files containing the data to be restored.
  - `<USERNAME>` is the Sybase login name.
  - `<SERVER>` is the Sybase Server for the Data Pool database (e.g., x4dbl03\_svr).
  - `<DBNAME>` is the name of the Data Pool database (e.g., DataPool OPS).
  - The script displays a prompt for entry of the password for the Sybase login.

**NOTE:** The step that follows may require input from the Database Administrator.

- 4 Type `<password>` and then press the **Return/Enter** key.
    - The script runs and the Archive Access Statistics Utility log file `DIDbRestoreAccessStat.log` records errors, warnings, and information about utility events. The log is written to the directory `/usr/ecs/<MODE>/CUSTOM/logs`.
- 

### 14.11.4 Running the Batch Insert Utility

The Batch Insert Utility is being replaced by Publish Utility. Please refer to the related section (14.11.12).

#### 14.11.4.1 Running the Batch Insert Utility

---

The Batch Insert Utility is being replaced by Publish Utility. Please refer to the related section (14.11.12).

#### 14.11.5 Running the Most Recent Data Pool Inserts Utility

The **Most Recent Data Pool Insert Utility (EcDIMostRecentInsert)** lists the most recent additions to the Data Pool. The output of the utility is a set of files that a user could download and quickly inspect to identify recent additions to the Data Pool.

The utility takes in a date command-line parameter indicating the day of interest to the user. Files inserted into Data pool on the specified day are subsequently listed in the output files. If no date is provided, the utility uses the preceding day as a default with a time range of midnight to midnight.

The Most Recent Data Pool Insert Utility normally runs as a cron job. However, if it is necessary to run the utility from the command line it is possible to do so.

The procedure for running the Most Recent Data Pool Insert Utility is based on the following assumptions:

- Data Pool database server is running.
- Data Pool database is available.
- Stored procedures are present.

##### 14.11.5.1 Running the Most Recent Data Pool Inserts Utility

---

- 1 Log in at the host where the Most Recent Data Pool Inserts Utility is installed (e.g. x4dpl01).
- 2 Type `cd /usr/ecs/<MODE>/CUSTOM/utilities`, then press the **Return/Enter** key.
- 3 Type `EcDIMostRecentInsert.pl <MODE> -insertDate <YYYY/MM/DD>` and then press the **Return/Enter** key.
  - `<MODE>` is the mode in which the utility is being executed (e.g., OPS, TS1, or TS2).
  - `-insertDate` is an optional parameter specifying date of user's interest. If the date parameter is not specified, the preceding day's date is used as the default value.
  - For example, if today were July 11, 2005, the following command: `EcDIMostRecentInsert.pl OPS` would generate files concerning additions to the Data Pool between midnight July 9, 2005 and midnight July 10, 2005.
  - The Most Recent Data Pool Inserts Utility runs and generates a set of files:
  - One file, named `DPRRecentInserts_<YYYYMMDD>`, is located at the top-level Data Pool directory. It contains distinct GroupID, ShortName, and VersionID. For example, the file `DPRRecentInserts_20051102` in directory `/datapool/OPS/user` might

contain the following types of entries:

```
START_FILE: Entries:: 7
GROUP_ID SHORT_NAME VERSION_ID
MOAT MOD02QKM 077
MOAT MYD02OBC 077
MOAT MYD02QKM 077
MOAT MYD35_L2 077
TEST3 MOD35_L2 077
END_FILE: Written 7
```

- There is a file in each of the collection-level directories named `DPRecentInserts_<ShortName>_<VersionID>_<YYYYMMDD>`. The files contain `SHORTNAME`, `VERSION_ID`, and `RELATIVE_PATH`. For example, the file `DPRecentInserts_MYD35_L2_077_20051102` in directory `/datapool/OPS/user/MOAT/MYD35_L2.077` might contain the following types of entries:

```
START_FILE: FileEntries:: 616
SHORTNAME VERSION_ID RELATIVE_PATH
MYD35_L2 077 /MYD35_L2.077/2001.07.01/labtest_2019099138
MYD35_L2 077 /MYD35_L2.077/2001.07.01/labtest_2019099140
MYD35_L2 077 /MYD35_L2.077/2001.07.01/labtest_2019099237
MYD35_L2 077 /MYD35_L2.077/2001.07.01/labtest_2019099247
MYD35_L2 077 /MYD35_L2.077/2001.07.01/labtest_2019099263
[...]
```

- If it is unable to create a file at the top-level Data Pool directory, the Most Recent Data Pool Inserts Utility shuts down and logs an error message.
- If it is unable to create a file at a collection level directory, the Most Recent Data Pool Inserts Utility logs an error message and continues processing other valid directories.
- When the Most Recent Data Pool Inserts Utility runs, it creates a log file, `EcDIRMostRecentInsert.log`, which records errors, warnings, and information about utility events.
- The Most Recent Data Pool Inserts Utility does not perform automatic recovery operations. If there is an execution failure as a result of database server or system shut down, rerun the script. This will create a new set of files (overwriting previous ones) listing additions to Data Pool for the specified insert date.

---

### 14.11.6 Running the Data Pool Collection-to-Group Remapping Utility

The **Data Pool Collection-to-Group Remapping Utility (EcDIRemap)** is a command-line utility interface that is used for reassigning a Data Pool collection to a collection group other than the one to which it was originally assigned.

The procedure for running the Data Pool Collection-to-Group Remapping Utility is based on the following assumptions:

- “Insert Enabled Flag” for the source collection has been turned off using the Data Pool Maintenance GUI
- The group to which the user is mapping the collection already exists in the Data Pool database.
- The group to which the user is mapping the collection is not the BRWS (browse) group.
- The collection to be remapped is not the Browse (Browse.001) collection.
- Data Pool database server is running.
- Data Pool database is available.
- Stored procedures are present.

There are several assumptions expected of the Data Pool Collection-to-Group Remapping Utility. The utility expects the existence of the collection in the Data Pool to which the user is mapping; it assumes that the browse collection is always located in the group “BRWS”; it also assumes the stored procedures are present. The Group Mapping utility runs only if the Data Pool database server is running and if the database is available.

#### 14.11.6.1 Running the Data Pool Collection-to-Group Remapping Utility

---

- 1 Ensure that the “Insert Enabled Flag” for the source collection has been turned off.
  - For detailed instructions refer to the **Modify an ECS Collection Using the DPM GUI** procedure (previous section of this lesson).
- 2 Log in at the host where the Data Pool Collection-to-Group Remapping Utility is installed (e.g., x4dpl01).
- 3 Type `cd /usr/ecs/<MODE>/CUSTOM/utilities`, then press the **Return/Enter** key.
- 4 Type the following:  
**EcDIRemap.pl <MODE> -esdt <name> -version <version> -oldgrp <old group> -newgrp <new group>** and then press the **Return/Enter** key.
  - <MODE> is the mode in which the utility is being executed (e.g., OPS, TS1, or TS2).
  - <name> is the name of the source collection being remapped.
  - <version> is the version of the source collection version being remapped.
  - <old group> is the name of the collection group name that currently contains the collection.
  - <new group> is the name of the collection group to which the source collection is being remapped.

Example: To remap collection MOD29, Version 4, (i.e., MOD29.004) from collection group MOST to collection group MOSS, enter the following:

```
EcDIRemap.pl OPS -esdt MOD29 -version 4 -oldgrp MOST -newgrp
MOSS
```

- The Data Pool database inventory would be updated to reflect the new location of the files.
  - The Data Pool Collection-to-Group Remapping Utility runs and the log file, **EcDIRemap.log**, records errors, warnings, and information about utility events.
  - The Data Pool Collection-to-Group Remapping Utility is able to recover from aborted runs by using the DIRcoveryParameters table to checkpoint its progress. In the event of an aborted run, the utility reads the recovery parameters table to determine at which point the utility left off when it aborted. This ensures that remappings that were taking place prior to the abort finish correctly. After recovery processing takes place, the utility processes the current run by acting on the latest input parameters.
- 

#### **14.11.7 Running the Data Pool Move Collections Utility**

The Move Collections Utility provides the EED Operations Staff with a command-line interface to move collections from one file system to another. The utility requires command-line parameters that specify the collection (shortname and version id) to be moved and the source and target file system path. The utility also supports a verbose and debug option. This verbose option allows for enabling or suppressing detailed information displayed to the screen and log for the operator. The default is non-verbose, which allows an operator to run the utility as a background process. Fault recovery is also supported, allowing completion of a partially moved collection due to a database server fault or an operating system error.

The utility relies on the fact that symbolic links will be set from the collection's old filesystem to its new filesystem. For example, before a move, a collection might be located here: /datapool/OPS/user/FS1/MOAT/AIRABRAD.007. After invoking the utility with a target filesystem of FS2, it will be moved to /datapool/OPS/user/FS2/MOAT/AIRABRAD.007 with a symbolic link from its old location, i.e. /datapool/OPS/user/FS1/MOAT/AIRABRAD.007 → /datapool/OPS/user/FS2/MOAT/AIRABRAD.007

These links will be persistent so that the data can still be retrieved without changing a URL. However the URLs for all the granules in the collection that were moved are exported to ECHO. The file system move is implemented as a copy operation to the new collection directory location, followed by removal of the old collection directory and its contents. Between the first copy operations and the cleanup/removal operations, the AmCollection table is updated with the target file system label. The utility then establishes a link to the new location at the collection level. As a result, existing URLs will not be invalidated by the move and no updated URLs need to be exported to ECHO. However, existing URLs and file pointers will be invalid from the time when the utility starts deleting the existing directories until the time the link is established. During this time:

- A Data Pool ftp user clicking on a URL might experience a temporary error when trying to access files and directories associated with the moving collection. File transfers that are already in progress when deletion begins should complete normally.
- FTP Pull users could experience similar temporary problems when they try to access links in FTP Pull directories that were established by the OMS and that point to granules in the moving collection.

In addition, the following errors may occur during a collection move:

- During the time a collection is being moved, the Data Pool Web GUI will return an error if a user tries to access the collection via a bookmark. It will flag the collection and not display it as an active link on the collection drill down web page, temporarily preventing drill down access to the collection.
- The Data Pool insert service will look up the collection path in the Data Pool database during the insert process. The collection path is updated once the initial copy phase is completed and before deletion/copy phase. Any Data Pool insert processes that looked up the file system label BEFORE it was updated will insert their granules into the old directory location. The Data Pool InsertActionQueue table should be checked around the file system label update time (found in the MoveCollection<shortname> log) to determine if publications occurred successfully.

The OMS looks up granule file locations immediately before performing an FTP Push operation. If the lookup occurs just before the collection information in the Data Pool database is updated, but the copy operation starts after the file was deleted, the FTP Push operation will fail and cause an operator intervention. Since the interval of time between file location look up and ftp push start is small, the chances for that occurring are very small. The operator would need to resubmit the request, and since the directory entry will now have been updated, the ftp push operation will succeed. If the above impacts are not acceptable, operators can suspend inserts and web access for the original file system by marking it as "unavailable" in the DPM GUI. This would also halt staging operations for that file system in OMS. To prevent errors, operators would need to verify before activating an order that it does not reference granules from the collection that is being moved. Operations will need to use a different mechanism to alert FTP users of the unavailability; and to prevent access, operations would need to take other measures, such as changing the directory permissions.

There are various command line parameters that are used in combination with each other. Table 14.11-2 provides a description of these parameters.

**Table 14.11-2. Command Line Parameters**

| Parameter Name                 | Description                                                                                                                                                                                                                                                                                                                       |
|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <mode>                         | An input parameter that specifies the mode of operation. This must be the first parameter passed, and it must be a valid, existing Data Pool mode with a format like OPS or TS1. This parameter is mandatory.<br><b>Note:</b> The user will be prompted if the utility is run in OPS mode to prevent any accidental loss of data. |
| -verbose                       | Directs the utility to run using verbose option. Some information will be displayed to the screen and detailed information will be written to the utility's log. Default is nonverbose. (See Note)                                                                                                                                |
| -shortname <shortname>         | An input parameter that specifies the shortname of the collection to be moved. This parameter is mandatory.                                                                                                                                                                                                                       |
| -versionid <versionid>         | An input parameter that specifies the version identifier of the collection to be moved. Do not specify leading zeros. This parameter is mandatory.                                                                                                                                                                                |
| -sourcefs <file system label>  | An input parameter that specifies the source file system label (i.e. FS1 or FS2) from which the collection is being moved. Note that all Data Pool file systems must be mounted under the Data Pool root (e.g. (/datapool/OPS/user)). This parameter is mandatory.                                                                |
| -targetfs <file system label > | An input parameter that specifies the relative target file system path to which the collection is being moved. Note that all Data Pool file systems must be mounted under the Data Pool root (e.g. (/datapool/OPS/user)). This parameter is mandatory.                                                                            |
| -debug                         | This option directs the utility to print out other debug information. Default is no debug.                                                                                                                                                                                                                                        |

There is no required ordered sequence of the parameters except for the parameter <MODE>. This must be first parameter or a fatal error will be returned. The combination of these remaining inputs must be valid. A command line input error results in a 'usage' syntax display, and in most cases will also explain why the input was incorrect.

### Configuration File Format – EcDIMoveCollection.CFG

The "config" file contains vital details about how to connect to the Sybase database. Without this file, the utility can not run. The config file must be a single-entry plain text ASCII file, which has the following format:

```

SYB_USER = <string>
SYB_SQL_SERVER = <string>
SYB_DBNAME = <string>
PGM_ID = <string>
NUM_DB_RETRIES=<integer>
DB_SLEEP_SEC=<integer>
NUM_DELETE_RETRIES=<integer>
DELETE_SLEEP_SEC=<integer>

```

See Table 14.11-3 for a breakdown of configuration file parameters.

**Table 14.11-3. Configuration File Parameters**

| Parameter Name     | Description                                                                                                                                                                                                                                                                                                                   |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SYB_USER           | The user name for the Sybase connection.                                                                                                                                                                                                                                                                                      |
| SYB_SQL_SERVER     | The name of the SQL server for this Sybase connection.                                                                                                                                                                                                                                                                        |
| SYB_DBNAME         | The name of the database you intend to connect to                                                                                                                                                                                                                                                                             |
| PGM_ID             | Program ID used for connecting to the Data Pool database. The value of this parameter must be set to 10000022 for this program.                                                                                                                                                                                               |
| NUM_DB_RETRIES     | The number of times the utility will attempt to connect to the database before exiting. The recommended default is 5.                                                                                                                                                                                                         |
| DB_SLEEP_SEC       | The number of seconds the utility will wait ('sleep') between connection attempts. The recommended default is 10.                                                                                                                                                                                                             |
| NUM_DELETE_RETRIES | The number of times the utility will rescan the old collection directory prior to deleting it. If the delete fails, it is most likely because the directory is not empty because some granules were inserted after the move started. The repeated rescanning for these files handles this case. The recommended default is 5. |
| DELETE_SLEEP_SEC   | The number of seconds the utility will wait ('sleep') between old collection directory rescans/deletes. The recommended default is 10.                                                                                                                                                                                        |

#### 14.11.7.1 Running the Data Pool Move Collections Utility

- 1 Log in at the host where the Data Pool Move Collections Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Data Pool Move Collections Utility, type the following:  
**cd /usr/ecs/<MODE>/CUSTOM/utilities**, then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the Batch Insert Utility, in the form of the following.

```
EcDlMoveCollection.pl <mode> -shortname <shortname>
-versionid <versionid> -sourcefs <file system label>
-targetfs <file system label> -verbose -debug
```

Data Pool Move Collections Utility usage examples

- **EcDlMoveCollection.pl <mode> -shortname MODVOLC**  
**-versionid 1 -sourcefs FS1 -targetfs FS2 -verbose**

Moves the files, browse links, and inventory information for the collection **MODVOLC.001** from its current directory structure in file system FS1 to the new filesystem FS2. The collection will be moved from /datapool/<mode>/user/FS1/MOAT to /datapool/<mode>/user/FS2/MOAT. The utility will be run using the -verbose option, which displays information to the screen and to the log.

- `EcDlMoveCollection.pl <mode> -shortname MODVOLC  
-versionid 1 -sourcefs FS1 -targetfs FS2`

Same as 1) but in non-verbose mode. No output to the screen and less detail in the log.

---

#### 14.11.8 Running the Data Pool Hidden Scrambler Utility in Rename Mode

The Data Pool Insert Service stores granules that are staged to the Data Pool for ordering purposes only in separate directories whose contents are not visible to anonymous ftp users. Order-only granules were accessible by the general public.

When an order-only granule is subsequently inserted via a “normal” Data Pool insert, it becomes a normal Data Pool granule, and the Data Pool Insert Service moves it from the hidden directory to the appropriate place in the public directory structure. Of course, such transfers are not allowed for billable or restricted granules. DAACs should not and generally do not insert granules from billable collections into the Data Pool, and the Data Pool Insert Service performs various checks (including examination of the DFA flag) to prevent the insert of granules that are flagged as “Hidden” or “Restricted” in ECS.

To decouple Data Pool and OMS file references, ordered granules are always represented in the hidden directory structure, either by files (if the granule is not in the public Data Pool) or by links pointing to the public files (if the granule is in the public Data Pool). Whenever a granule is converted from an order-only to a public Data Pool granule, its files are moved and links are left behind in the order only directories. Although a “metadata only” granule would be considered public, its science file would remain inaccessible (i.e., it would never be in the public Data Pool).

The hidden directory structure is below the FTP root because of the need to support FTP Pull access to ordered data. To hide the directories, the FTP root contains a directory that serves as the root for the hidden directory structure. (The directory is directly below the file system level, i.e., at the level of the Data Pool collection groups.) While it is impossible to hide that directory from view, it and all directories below it are configured in such a manner that their contents cannot be listed via ftp, in effect hiding all lower-level directories as well as their contents from public view.

The hidden directory structure mimics the public Data Pool directory structure (i.e., it is organized by collection group and collection); however, the hidden directory structure uses encrypted directory names so the pathnames cannot be guessed, preventing anonymous FTP users from switching into a hidden directory via the `cd` command. The Data Pool Hidden Scrambler Utility (`EcDIHiddenScrambler.pl`) allows the DAAC to re-encrypt directory names during system maintenance periods (i.e., during Data Pool down time), either on a regular basis or when intrusion is suspected.

Other Data Pool utilities (e.g., Data Pool collection move and re-map utilities) have been changed to the extent that they rely on the directory naming conventions so they can deal with granules in the hidden directory structure.

OMS takes responsibility for removing order-only data from the Data Pool when they are no longer needed. However, the responsibility for cleaning up the public Data Pool remains with the Data Pool Cleanup Utility.

It is essential that the names of the hidden directory names do not become public knowledge. An external user could use knowledge of directory names and clever guessing of file names to download from them via anonymous FTP. The Data Pool cannot prevent this because it is necessary to allow FTP Pull download from the directories via anonymous FTP. However, the Data Pool log analyzer will detect any attempt to access the hidden directories directly and will send an e-mail message to a DAAC-configured address to report security breaches. When that occurs, the DAAC should shut down FTP access to the Data Pool as soon as possible and run the Data Pool Hidden Scrambler Utility, which generates a new set of hidden directory names and updates the existing data holdings.

The Data Pool Collection-to-Group Remapping Utility will move the hidden collection directories when it moves the public collection directories. The Data Pool Move Collections Utility may also move the hidden directories for a collection depending on whether the order-only files are in the same file system as the public collection or are in a designated file system of their own.

The Data Pool Hidden Scrambler Utility (EcDIHiddenScrambler.pl) can be run in either of the following two modes:

- Transition.
- Rename.

In transition mode the utility generates hidden directory names and corresponding database entries for every collection defined for Data Pool in the affected operating mode. The transition mode can be used while Data Pool is up. The utility should be run in transition mode only once; i.e., the first time the utility is run in any given operating mode. Because transition mode is not used during normal operation, it is not described in any detail in this section.

In rename mode the utility re-encrypts all of the hidden directory names. This involves updates to the directory in the file system and to the database. Links from the FTP Pull area (and elsewhere) are preserved. Re-encryption must be done during DAAC downtime only.

If the Data Pool Hidden Scrambler Utility is interrupted during execution, upon restart it detects failures from the previous run and continues processing the directories and files that were left unprocessed in the previous run. The operator is given no choice as to recovery. Recovery proceeds so that the Data Pool inventory and disk files will not be left in a corrupted state.

The procedure for running the Data Pool Hidden Scrambler Utility in rename mode starts with the assumption that the Data Pool is down for maintenance (no orders being processed, no external access to the Data Pool for downloading data, etc.).

**THE DATA POOL HIDDEN SCRAMBLER UTILITY SHOULD BE RUN IN TRANSITION MODE ONLY ONCE; I.E., THE FIRST TIME THE UTILITY IS RUN IN**

**ANY GIVEN OPERATING MODE. IN NORMAL OPERATIONS, THE DATA POOL HIDDEN SCRAMBLER UTILITY IS RUN IN RENAME MODE.**

### **14.11.8.1 Running the Data Pool Hidden Scrambler Utility in Rename Mode**

---

**NOTE:** In normal operations, the Data Pool Hidden Scrambler Utility is run in rename mode whenever hidden directory intrusion is detected/suspected. In addition, it is recommended that the Data Pool Hidden Scrambler Utility be run in rename mode on a scheduled basis (e.g., monthly) at the DAAC's discretion.

- 1** Log in at the machine where the Data Pool Hidden Scrambler Utility is installed (e.g., x4dpl01).
  - The script must be run from a user account with privileges to rename directories on the Data Pool.
- 2** Type `cd /usr/ecs/<MODE>/CUSTOM/utilities` then press **Return/Enter**.
  - Change directory to the directory containing the Data Pool Hidden Scrambler Utility script (e.g., EcDIHiddenScramblerDataPool.pl).
  - <MODE> will most likely be one of the following operating modes:
    - OPS (for normal operation).
    - TS1 (for SSI&T).
    - TS2 (new version checkout).
  - Note that the separate subdirectories under /usr/ecs apply to different operating modes.
- 3** To perform a “rename” run, at the UNIX prompt enter one of the following:
  - `EcDIHiddenScramblerDataPool.pl <MODE>`
  - OR
  - `EcDIHiddenScramblerDataPool.pl <MODE> -collgroup <collgroup>`
  - OR
  - `EcDIHiddenScramblerDataPool.pl <MODE> -shortname <shortname> -versionid <versionid>`
  - <collgroup> is a particular Data Pool collection group with collection directories to be renamed using the Hidden Scrambler Utility. If the –collgroup parameter is specified, the –shortname and –versionid parameters may not be used. If not all directories for collections within a collection group are to be renamed, run the Hidden Scrambler Utility using the –shortname and –versionid parameters to rename the directory for each affected collection.
  - <shortname> is the name of a particular Data Pool collection, the directory for which is to be renamed using the Hidden Scrambler Utility. If the –shortname parameter is specified, the –versionid parameter must be specified too. If the –shortname parameter is specified, the –collgroup parameter may not be used.

- *<versionid>* is the version ID of a particular Data Pool collection, the directory for which is to be renamed using the Hidden Scrambler Utility. If the *-versionid* parameter is specified, the *-shortname* parameter must be specified too. If the *-versionid* parameter is specified, the *-collgroup* parameter may not be used.
- The following examples show valid command-line entries for a “rename” run of the Hidden Scrambler Utility:

**EcDIHiddenScramblerDataPool.pl OPS**

- The Hidden Scrambler Utility performs rename processing (re-encrypts the hidden directory names) for all collection groups and all collections in the Data Pool in OPS mode.

**EcDIHiddenScramblerDataPool.pl OPS -collgroup MOAT**

- The Hidden Scrambler Utility performs rename processing (re-encrypts the hidden directory names) in OPS mode for the MOAT collection group, including all collections in the MOAT collection group.

**EcDIHiddenScramblerDataPool.pl OPS -shortname AST\_L1B -versionid 3**

- The Hidden Scrambler Utility performs rename processing (re-encrypts the hidden directory names) in OPS mode for the AST\_L1B.003 collection. [Note that the hidden directory name of the corresponding collection group (ASTT) would not be re-encrypted.]

- If applicable, usage errors (e.g., failure to specify a mode) are displayed on the terminal screen.
- The Data Pool Hidden Scrambler Utility records events and errors in the **EcDIHiddenScrambler.log** (in the */usr/ecs/<MODE>/CUSTOM/logs* directory). If the log file exists already, the new information is automatically appended. If there is no existing log file named *EcDIHiddenScrambler.log*, the utility creates a new log file with that name.
- The Data Pool Hidden Scrambler Utility provides a capability to recover from interruptions caused by situations such as the system faults or database errors that leave all or some of the directories unprocessed. The utility detects such failure upon the next run and continues processing the directories and files that were left unprocessed in the previous run. The operator is given no choice as to recovery. Recovery proceeds so that the Data Pool inventory and disk files will not be left in a corrupted state.

### 14.11.9 Running the Data Pool Cleanup Orphan/Phantom Validation

*EcDICleanupFilesOnDisk.pl*, a new utility, provides a mechanism for the ECS Operator to perform Data Pool orphan/phantom validation, and to remove the orphaned data from the Data Pool Database inventory and the Data Pool disks if desired.

EcDICleanupFilesOnDisk.pl replaces EcDICleanupDataPool.pl for orphan/phantom validation, as well as orphan file cleanup. Several enhancements have been made as part of the Release 7.22 as listed below:

- a. The utility will validate the orphan/phantom files but not remove them from the data pool disks.
- b. The utility will remove an orphaned file from the Data Pool inventory and disk if the –fix option is specified.
- c. The utility will validate an orphaned browse granule that does not have any cross-references with any public granules (i.e., a granule whose isOrderOnly in (NULL, ‘B’)) in the Inventory Catalog.
- d. The utility will remove the orphaned browse granule, and its files from the Data Pool inventory and Data Pool disks if the–fix option is specified.
- e. If the max orphan age parameter on the command line is less than three days, override the parameter with the three-day value; a message as below will be displayed on the screen and be logged in the log file.  
**“The minimum possible value for maxOrphanAge is 3 days. Files modified less than 3 days ago will not be considered not in database”**
- f. The utility will skip collection groups not found in the DICollectionGroup table, and will output and log invalid collection groups as:  
**“Invalid collection groupId [collection group name]”**  
The Data Pool Inventory Validation shall skip validation for collections that reside on file systems that are currently marked as unavailable or are suspended if the validation requires access to that file system, and log the collections that are skipped because of this.
- g. If all collection groups specified are invalid, then the utility will terminate with an exit code of 1. The following error message will be output and logged:  
**“No collection groups are specified or all groups are invalid”**
- h. The utility will skip and log a collection whose file system is unavailable when the script starts. The error message is as:  
**“File system [file System] is unavailable for collection [collection name]”**
- i. The exit code is specified for the following situations:
  - 0 - successful validation with no discrepancies
  - 1 - failed to successfully run utility (internal error)
  - 2 - successful validation with discrepancies

Table 14.11-4 provides a description of these parameters.

**Table 14.11-4. Command Line Parameters**

| Parameter Name | Required | Description                                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| mode           | Yes      | Specifies the mode of operation. The MODE parameter is mandatory as the first parameter.                                                                                                                                                                                                                                                                                                                     |
| collgroup      | No       | Allows user to specify a list of collection groups to perform validation on.                                                                                                                                                                                                                                                                                                                                 |
| fix            | No       | Allows user to delete the orphaned files.                                                                                                                                                                                                                                                                                                                                                                    |
| maxFileAge     | No       | Specifies how old the file must be before deleting it. The minimum possible value for maxFileAge is 3 days. Files modified less than 3 days ago will not be considered not in database.                                                                                                                                                                                                                      |
| outputDir      | No       | By default the output files are written to /usr/ecs/<MODE>/CUSTOM/data/DPL/Validation. If -outputDir option is specified, the output file should locate in the /usr/ecs/<MODE>/CUSTOM/data/DPL/Validation/<outputDir>, the value of -outputDir should be <outputDir>. If the directory <outputDir> does not exist, then the utility creates it.                                                              |
| debug          | No       | <p>a. log more information for error checking.</p> <p>b. Create two files for a give collection, one with files list found from database , the other with the files list found from disk, all locate in /usr/ecs/&lt;MODE&gt;/CUSTOM/temp/DPL</p> <p>Here are examples of those two files:</p> <p>FilesInDb.GLA04.028aLfGVsJi.6041.20081126143245</p> <p>FilesOnDisk.GLA0786XInYQGSb.3187.20081126142747</p> |

## EcDlCleanupFilesOnDisk Configuration File

The utility uses a configuration file, EcDlCleanupDataPool.CFG, located in the /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table .14.11-5 describes the configuration parameters.

**Table 14.11-5. Configuration Parameters**

| Parameter Name              | Description                                                                                                                                                                                                       |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SYB_USER                    | Sybase login name for the user of the Data Pool database.                                                                                                                                                         |
| SYB_SQL_SERVER              | Name of Sybase SQL Server hosting Data Pool database.                                                                                                                                                             |
| SYB_DBNAME                  | Name of Data Pool database.                                                                                                                                                                                       |
| PGM_ID                      | Program identifier used as seed to generate database password.                                                                                                                                                    |
| DEFAULT_LIMIT               | Default priority limit if a limit (-limit) is not provided via command line.                                                                                                                                      |
| NUM_RETRIES                 | Number of times database operation will be attempted.                                                                                                                                                             |
| SLEEP_SEC                   | Number of seconds between retries.                                                                                                                                                                                |
| MAX_ORPHAN_AGE              | Maximum age in days in qualifying a file as an orphan. A file must have an age greater than or equal to this value in order to be considered as an orphaned file. The parameter value must be 10 days or greater. |
| URL_EXPORT_RETENTION_PERIOD | The maximum age in days that a URL_EXPORT will be retained. The suggested value for this field is "30"                                                                                                            |
| ORDER_OUTPUT_COLLECTIONS    | A list of collection groups that should not be processed during orphan validation. Files in these directories will be ignored when determining orphans. The suggested value for this field is "OUTPUTS BRWS"      |
| MAX_ORDER_AGE               | The number of days for which a file in the order only directory will not be counted as an orphan. Suggested value for this field is 15.                                                                           |

### 14.11.9.1 Running the Data Pool Cleanup Orphan/Phantom Validation

---

- 1 Login at the machine where the Data Pool Cleanup Orphan/Phantom Validation Utility is installed (e.g., x4dp101).
  - The operator who is executing the script must have privileges for removing science, metadata, and browse files from the Data Pool disks.
- 2 To change to the directory for starting the Data Pool Cleanup Orphan/Phantom Validation Utility, type `cd /usr/ecs/<MODE>/CUSTOM/utilities` and then press the **Return/Enter** key.
  - The working directory is changed to `/usr/ecs/<MODE>/CUSTOM/utilities`.
- 3 At the prompt, type the command to start the Data Pool Cleanup Orphan/Phantom Validation Utility, in the form of the following.

**EcDlCleanupFilesOnDisk.pl <mode >**

|                                               |                    |
|-----------------------------------------------|--------------------|
| <b>[-collgroup &lt;groupList&gt;]</b>         | <b>-- optional</b> |
| <b>[-maxFileAge &lt;age in # of days&gt;]</b> | <b>-- optional</b> |
| <b>[-outputDir &lt;outputDir&gt;]</b>         | <b>-- optional</b> |
| <b>[-fix]</b>                                 | <b>-- optional</b> |
| <b>[-debug]</b>                               | <b>-- optional</b> |

The <mode> parameter is mandatory as the first parameter.

Data Pool Cleanup Orphan/Phantom Validation Utility usage examples

- For a "**validation only**" run:  
**EcDlCleanupFilesOnDisk.pl TS1**  
- for all collection groups, maximum file age will be 3 days, output files will be written to the directory `/usr/ecs/<MODE>/CUSTOM/data/DPL/Validation/`
- EcDlCleanupFilesOnDisk.pl TS1 – collgroup “MOST BRWS” -maxFileAge 5**  
- for collection groups “MOST BRWS” only, maximum file age will be 5 days, output files will be written to the directory `/usr/ecs/<MODE>/CUSTOM/data/DPL/Validation/`
- EcDlCleanupFilesOnDisk.pl TS1 -outputDir mytest**  
- for all collection groups, maximum file age will be 3 days, output files will be written to the directory `/usr/ecs/<MODE>/CUSTOM/data/DPL/Validation/mytest/`
- EcDlCleanupFilesOnDisk.pl TS1 -debug**  
- for all collection groups, maximum file age will be 3 days, output files will be written to the directory `/usr/ecs/<MODE>/CUSTOM/data/DPL/Validation/`, and two additional files will be written to `/usr/ecs/<MODE>/CUSTOM/temp/DPL/`

- For a "cleanup following validate" run:  
**EcDILCleanupFilesOnDisk.pl TS1 – collgroup “MOST BRWS” -maxFileAge 5 -outputDir mytest –fix**
- 

### 14.11.10 Running the Data Pool SoftLink Check Utility

EcDILinkCheck.ksh found a ‘broken’ symbolic link in the public Data Pool directories, i.e., a link whose target does not exist. Public Data Pool directories contain only Browse links, so the symbolic link would be a browse link.

Most likely, the link was supposed to be removed and that failed for some reason. Note, though, that the utility may include links in the report that were only temporarily incorrect. Therefore, DAAC staff should verify that the reported links are indeed broken and then remove broken links via UNIX command or with –fix option.

Table 14.11-6 provides a description of these Command Line Parameters.

**Table 14.11-6. Command Line Parameters**

| Parameter Name           | Required | Description                                            |
|--------------------------|----------|--------------------------------------------------------|
| DIRECTORY_TO_START CHECK | Yes      | Directory to start looking for ‘broken’ symbolic links |
| FIX                      | No       | Search and remove invalid links.                       |

#### 14.11.10.1 Running the Data Pool SoftLink Check Utility

---

- 1 Log in at the machine where the Data Pool SoftLink Check Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Data Pool SoftLink Check Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.
- 3 At the prompt, type the command to start the Data Pool SoftLink Check Utility, in the form of the following.

**EcDILinkCheck.ksh < DIRECTORY\_TO\_START\_CHECK > <-fix>**

Data Pool SoftLink Check Utility usage examples:

- **EcDILinkCheck.ksh /datapool/TS2/user/FS1/MOLA –fix**  
The utility will search for broken links under /datapool/TS2/user/FS1/MOLA directory, then remove the invalid links.  
If the directory /datapool/TS2/user/FS1/MOLA does not exist or the file system FS1 is down, the utility will exit with code 1.
- **EcDILinkCheck.ksh /datapool/TS2/user/FS1/MOLA**

The utility will search for broken links under /datapool/TS2/user/FS1/MOLA directory, and write results to an output file with a list of filename with full path to /usr/ecs/<mode>/logs/EcDILinkCheck.log

---

### 14.11.11 Running the Data Pool Online Archive Cleanup Utility

EcDlCleanupGranules.pl, a Data Pool Cleanup utility, provides a mechanism by which the ECS Operations Staff can remove granules and their associated metadata and browse files from the Data Pool disks and the corresponding Data Pool database inventory.

Qualification of a granule for cleanup is based on following criteria:

- a. The utility will remove non-ECS granules only when DAAC staff explicitly requests their removal.
- b. The utility will handle the deletion of non-ECS granuleIds specified on the command line or in a file.
- c. The utility will transparently cleanup the Data Pool when collections reside on different file systems.
- d. The utility will postpone the cleanup of a granule that is currently in use by the OMS CI, until such time as OMS relinquishes use of that granule.
- e. The utility will only remove non-ECS granules by expiration date/time and retention priority when the `-expired` option is specified.
- f. The utility will remove a collection directory if it is empty, as well as any link to the collection directory that was established during a collection move operation.
- g. The utility will skip the cleanup of granules belong to collections on file systems that are marked unavailable. They will be retried next run.
- h. The utility will export granules that will be deleted to ECHO at the start of program The utility will support a batch file delete if the `-batchsize` option is specified.
- i. By default, the utility will handle recovery of unprocessed granules that were left in tables `DIbcpGransToDelete` and `DIBatchGransToDelete` from the failure of a previous program execution. `EcDlCleanupGranules.pl` provides a `norecovery` option.
- j. The utility will check the syntax of the command line parameters and display the error and the correct command line syntax if the command line parameters fail the syntax check.
- k. The utility will check command line input parameters, if there are any conflicting or invalid parameters detected, the utility will exit with an exit code of 1 and display the error and the correct command line input parameters.
- l. When performing a cleanup on ECS granules the utility will update the Inventory Catalog to indicate the granule is stored in the “hidden” Online Archive, and remove all files for the granule from the Online Archive. This usage provides a mechanism for the Operator to restore an ECS granule to a “known state” after a publication failure. When running `EcDlCleanupGranules` with ECS granules, it is also necessary to run the

EcDIRestoreOlaFromTape utility to copy the files back to the Online Archive from the tape Archive.

- m. The exit code is specified for the following situations:
  - 0 - successful cleanup with no discrepancies
  - 1 - failed to successfully run cleanup (internal error)
  - 2 - successful cleanup with discrepancies

Table 14.11-7 provides a description of Command Line Parameters.

**Table 14.11-7. Command Line Parameters**

| Parameter Name | Required | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|----------------|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| File           | No       | File name with a list of DPL granulesIds as input                                                                                                                                                                                                                                                                                                                                                                                                                |
| Grans          | No       | a list of DPL granuleIds from the command line                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Offset         | No       | Specifies hours before (negative) or after (positive) midnight of the previous day from which to delete. Defaults to zero.<br>(Some examples:<br><b>-offset 5</b> would delete all granules which had expired as of 5 AM of the current day;<br><b>-offset -5</b> would delete all granules which had expired as of 7 PM yesterday<br><b>-offset 72</b> would delete all granules which will be expiring in 72 hours measured from the previous day's midnight). |
| Limit          | No       | Specifies limiting value used for determining which granules will be deleted. Will delete all granules with priority less than or equal to the specified limit. Must be within the range 1–255, 1 being the lowest priority and 255 being the highest priority. Defaults to value specified in configuration file.                                                                                                                                               |
| Theme          | No       | Specifies the name of a theme for which cleanup is to be performed. The Cleanup Utility will clean up non-ECS granules that would otherwise qualify for cleanup only if the granules are associated with that theme, and remove the granules entirely if they are not associated with any other theme, otherwise only remove the cross references with that theme. The theme name must be enclosed in quotes (").                                                |
| batchSize      | No       | Process cleanup by batch files. Recommend the batchSize to 100                                                                                                                                                                                                                                                                                                                                                                                                   |
| norecovery     | No       | Do not recover unprocessed granules that were left from the failure of a previous program execution. Also remove granules that were not deleted because they were on order or file system was unavailable                                                                                                                                                                                                                                                        |
| expired        | No       | Cleanup non-ECS granules by expiration date and retention priority                                                                                                                                                                                                                                                                                                                                                                                               |

### EcDICleanupGranules Configuration File

The utility uses a configuration file, EcDICleanupDataPool.CFG, located in the /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a

PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table 14.11-8 provides a description of Configuration Parameters.

**Table 14.11-8. Configuration Parameters**

| Parameter Name              | Description                                                                                                                                                                                                      |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SYB_USER                    | Sybase login name for the user of the Data Pool database.                                                                                                                                                        |
| SYB_SQL_SERVER              | Name of Sybase SQL Server hosting Data Pool database.                                                                                                                                                            |
| SYB_DBNAME                  | Name of Data Pool database.                                                                                                                                                                                      |
| PGM_ID                      | Program identifier used as seed to generate database password.                                                                                                                                                   |
| DEFAULT_LIMIT               | Default priority limit if a limit (-limit) is not provided via command line.                                                                                                                                     |
| NUM_RETRIES                 | Number of times database operation will be attempted.                                                                                                                                                            |
| SLEEP_SEC                   | Number of seconds between retries.                                                                                                                                                                               |
| MAX_ORPHAN_AGE              | Maximum age in days in qualifying a file as an orphan. A file must have an age greater than or equal to this value in order to be considered as an orphaned file. The parameter value must be 3 days or greater. |
| URL_EXPORT_RETENTION_PERIOD | The maximum age in days that a URL_EXPORT will be retained. The suggested value for this field is "30"                                                                                                           |
| ORDER_OUTPUT_COLLECTIONS    | A list of collection groups that should not be processed during orphan validation. Files in these directories will be ignored when determining orphans. The suggested value for this field is "OUTPUTS BRWS"     |
| MAX_ORDER_AGE               | The number of days for which a file in the order only directory will not be counted as an orphan. Suggested value for this field is 15.                                                                          |

#### **14.11.11.1 Running the Data Pool Online Archive Cleanup Utility**

- 1 Log in at the machine where the Data Pool Online Archive Cleanup Utility is installed (e.g., x4dp101).
- 2 To change to the directory for starting the Data Pool Online Archive Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.
- 3 At the prompt, type the command to start the Data Pool Online Archive Cleanup Utility, in the form of the following.
  - Cleanup failed publications (restoring granules to a "known state" using a file containing a list of granules  
**EcDICleanupGranules.pl <mode> -file <inputFile>**  
**[-batchSize] <batch size> -- optional**

- Cleanup using a list of non-ECS granules  
**EcDlCleanupGranules.pl <mode> -grans <listOfGranuleIds>**  
**[-batchSize] <batch size> -- optional**
- Cleanup using an expiration date to delete non-ECS granules  
**EcDlCleanupGranules.pl <mode> -expired**  
**[-offset] <# of hours> -- optional**  
**[-limit] <priority limit> -- optional**  
**[-theme] <themeName> -- optional**  
**[-batchSize] <batch size> -- optional**

The MODE parameter is mandatory as the first parameter.

Data Pool Online Archive Cleanup Utility usage examples:

- **Cleanup using a file containing a list of granules:**  
 EcDlCleanupGranules.pl TS1 -file myfile -batchSize 100
  - **Cleanup using a list of granules**  
 EcDlCleanupGranules.pl TS1 -grans "30987 90876"
  - **Cleanup with -expired option**  
 EcDlCleanupGranules.pl TS1 -expired -offset 5 -limit 300 -theme "test"  
 -batchSize 100
  - **The utility will create a log file that will contain the following information:**  
 "Each file that has been removed"
- 

#### 14.11.12 Running the Data Pool Publish Utility

The DPL publish Utility is a command line tool that publishes specified granules from a file, command line or collection. It is primarily designed to publish granules that already exist in the Data Pool, but it can also be used to insert granules into the Data Pool from AIM. Note that the Publish Utility does not perform the insert and/or publication actions directly instead, it submits requests to the Data Pool Action Driver to perform the work on its behalf.

Table 14.11-9 provides a description of Command Line Parameters.

**Table 14.11-9. Command Line Parameters**

| Parameter Name                    | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| -file <input_file>                | Tells the publish utility to read the list of ECS ids of granules to be published from a file. <i>input_file</i> specifies the full path of the file.                                                                                                                                                                                                                                                                                                                                      |
| -g <id1>,<id2>...                 | Specifies the ECS ids of the granules to publish on the command line. Any number of granules may be provided (within any limits the shell places on command line length).                                                                                                                                                                                                                                                                                                                  |
| -collection <shortname.versionid> | Tells the Publish Utility to publish all granules belonging to a given collection.                                                                                                                                                                                                                                                                                                                                                                                                         |
| -maxnumconactions                 | Indicates the number of concurrent actions that may be submitted to the Data Pool Action Driver. This option can be use to limit the impact on existing operations. If not provided, it defaults to 5,000, which effectively assumes that it has exclusive use of the Action Driver.                                                                                                                                                                                                       |
| -register                         | Indicates that the Publish Utility should make sure that the given granules exist in the Data Pool. Any granule that does not exist in the Data Pool will be inserted (registered). No granules will be published (placed into the public Data Pool). Granules may be inserted into the Data Pool even if they are logically deleted, or marked as hidden (i.e. DeleteFromArchive = 'H'). Granules will not be inserted if it is marked as deleted from archive (DeleteFromArchive = 'Y'). |
| -publish                          | Indicates that the Publish Utility publish the given granules in the Data Pool. Only granules that already exist in the Data Pool will be published. Any granule that does not exist in the Data Pool will not be inserted. Granules that belong to a collection that is marked as not public (AllowPublishFlag='N'), or are logically deleted or hidden, will not be published. Note also, that older versions of a granule will not replace a newer version.                             |

### 14.11.12.1 Running the Data Pool Publish Utility

- 1 Log in at the machine where the Data Pool Publish Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Data Pool Publish Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.

At the prompt, type the command to start the Data Pool Publish Utility, in the form of the following.

```
EcDIPublishUtilityStart <MODE> -ecs [-file <file_name_path> | -collection
<ShortName.VersionId> | -g <ecsId1>[][,<ecsId2>] [-theme <themeName>] [-batchlabel
<batchLabel>] [-maxnumconactions <num>] [-register] [-publish] or $0 <MODE> -nonecs -
file <file_name_path> [-theme <themeName>] [-batchlabel <batchLabel>] [-
maxnumconactions <num>]
```

Data Pool Publish Utility usage examples

- ***EcDiPublishStart OPS -ecs -file /home/cmshared/granuleIds.txt***

Insert and publish granules for the granule ids contained in the specified file. The file contains one ECS granule id per line.

- ***EcDiPublishStart OPS -ecs -g 12345, 23456 -publish***

Publish the two hidden granules whose ECS ids are given.

- ***EcDiPublishStart OPS -ecs -collection MYD29P1D.004 -maxnumconactions 10***

Make sure all granules belonging to collection MYD29P1D version 4 are public in the Data Pool, limiting the number of concurrent Action Driver requests to 20. This is a low impact way to make sure a complete collection is public, but could take days to run to completion.

- ***EcDiPublishStart OPS -ecs -g 12345 -theme "test"***

Publish 1 ECS granule and establish the theme "test" to the granule.

- ***EcDiPublishStart OPS -nonecs -file /home/cmshared/nonecs\_xml.txt***

Publish nonecs granules which are specified in a list of xml files in "nonecs\_xml.txt".

---

### **14.11.13 Running the Data Pool UnPublish Utility**

The DPL Unpublish Utility is a command line tool that unpublishes specified granules from the Data Pool. Granules may be specified in a file, or by command line.

The Unpublish utility was developed for the on-line archive capability. It will:

- unpublish the specified science granules.
- remove associated browse granule if permitted.

The Unpublish utility can also be used to unpublish granules which are marked for deletion in the AIM Inventory database (DeleteEffectiveDate is set, or DeleteFromArchive flag is set to "Y" or "H"), for example, as would occur after a run of the Granule Deletion Utility.

Table 14.11-10 provides a description of Command Line Parameters.

**Table 14.11-10. Command Line Parameters**

| Parameter Name             | Description                                                                                                                                         |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| -file <input_file>         | The file which contains a list of DPL granule ids for unpublsh. Input_file specifies the full path and file name of the file.                       |
| -granules <id1>, <id2> ... | DPL granule ids for unpublsh.                                                                                                                       |
| -aim                       | Tells the unpublsh utility to unpublsh granules deleted from the AIM database. If this option is used, the --offset option should also be provided. |
| --offset <#days>           | Specifies the past number of days for which to find deleted AIM granules. This option is only valid in conjunction with the --aim option.           |
| -help                      | Display instructions to run the utility.                                                                                                            |

### 14.11.13.1 Running the Data Pool UnPublish Utility

1 Log in at the machine where the Data Pool Unpublish Utility is installed (e.g., x4dpl01).

2 To change to the directory for starting the Data Pool Unpublish Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.

- The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.

3 At the prompt, type the command to start the Data Pool Unpublish Utility, in the form of the following.

```
EcDIUnpublishStart.pl -mode <mode> [-file <input_file>] [-granules <id1>,<id2>...] [-aim --offset <#days>]
```

**EcDIUnpublishStart.pl -help** for instructions.

Data Pool Unpublish Utility usage examples:

- ***EcDIUnpublishStart.pl -mode OPS -file /home/cmshared/granuleIds.txt***

Unpublish public granules for the granuleIds contained in the specified file. The file contains one Data Pool granuleId per line.

- ***EcDIUnpublishStart.pl -mode OPS -granules 12345, 23456***

Unpublish public granules for the granuleIds specified in the command line, separated by “,”.

- ***EcDIUnpublishStart.pl -mode OPS -aim --offset 12***

Unpublish granules deleted from the AIM database since the current time – 12 days.

### 14.11.14 Running the Data Pool Inventory Validation Utility

The Online Archive Validation Tool provides the EED Operations Staff with a command-line interface to identify the discrepancies within the state of granules AIM Inventory database.

Table 14.11-11 provides a description of Command Line Parameters.

**Table 14.11-11. Command Line Parameters**

| Parameter Name   | Required | Description                                                                                                                                                                                                                                                                                                                                               |
|------------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mode             | Yes      | Specifies the mode of operation. The MODE parameter is mandatory as the first parameter                                                                                                                                                                                                                                                                   |
| outputDir        | No       | Specifies the relative path under base directory /usr/ecs/<MODE>/CUSTOM/data/DPL/Validation. Note: The base directory has to exist;<br>The relative directory (only one level down) will be created if it doesn't exist. This is where all the output files reside. If the relative path is not provided, the output files will go to the base directory. |
| suppressLDeleted | No       | When identifying granules that are missing in the DPL database, don't include the ones that have been logically deleted (deleteEffectiveDate is not null) in the AIM database.                                                                                                                                                                            |
| suppressDFAed    | No       | When identifying granules that are missing in the DPL database, don't include the ones that have been DFAed (DeleteFromArchive = "Y") in the AIM database.                                                                                                                                                                                                |

#### 14.11.14.1 Running the Data Pool Inventory Validation Utility

- 1 Log in at the machine where the Data Pool Inventory Validation Utility is installed (e.g., x4dpl01).
  - The operator who is executing the script must have privileges for removing science, metadata, and browse files from the Data Pool disks.
- 2 To change to the directory for starting the Data Pool Inventory Validation Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.
- 3 At the prompt, type the command to start the Data Pool Inventory Validation Utility, in the form of the following.

**EcDIInventoryValidationTool.pl <command line parameters>**

Data Pool Inventory Validation Utility usage examples:

- **EcDIInventoryValidationTool.pl DEV04 -outputDir inventory**

The relative directory under /usr/ecs/<MODE>/CUSTOM/data/DPL/Validation where all the outputs will reside. It will be created if it doesn't exist. If not provided, the output files will go to /usr/ecs/<MODE>/CUSTOM/data/DPL/Validation by default.

- **EcDIInventoryValidationTool.pl DEV04 -suppressLDeleted**

This checks the discrepancies in DPL and AIM databases and writes the output files in /usr/ecs/DEV04/CUSTOM/data/DPL/Validation/inventory directory. The output result won't include any granules that are logically deleted in AIM and missing in DPL.

- `EcDlInventoryValidationTool.pl DEV04 -suppressDFAed`

When identifying granules that are missing in the DPL database, don't include the ones that have been DFAed (DeleteFromArchive = "Y") in the AIM database.

---

#### **14.11.15 Running the Data Pool Checksum Verification Utility (Being replaced by Data Pool Checksum Verification Service(CVS))**

The DataPool Checksum Verification utility (DPCV) provides a mechanism by which the ECS Operations Staff can perform checksum verification for files in the Data Pool. It can be scheduled and run as a background process to proactively verify the integrity of files in the Data Pool. For example, the utility could be set up as a background process that would verify the checksum of a file every "Nth" month by specifying a checksum verification option based on time elapsed since the last time checksum was verified. The utility could also be run on-demand by the DAAC operator to verify checksum values for a particular set of files.

- The utility is capable of performing checksum verification by sampling files based on ESDT and insert date range, or elapsed time since the last time checksum was verified, or a given granule list.
- According to the sampling options specified, the utility scans the appropriate files and verify their checksum values.
- Upon successful checksum verification, the utility will update the time when checksum was verified for each file in the DataPool database.
- Upon detection of checksum verification failure after a configurable number of retry attempts, the utility will log detailed information about the failure which will include granule ID, ESDT, insert time, complete file path and file name, along with the checksum information -- including checksum type, checksum values (computed value vs. the corresponding value stored in database), the last time the file was checksummed, checksum origin (who performed the last checksum). This information will also be provided in a report produced by the utility at the end of a run.
- The verification report will also include statistical summary information including total number of files checked, number of files that failed checksum, percentage of files that failed checksum, categorized by ESDT.
- This utility will be designed such that the checksum verification can be throttled so it does not impact on-going daily operations.

Table 14.11-12 provides a description of Command Line Parameters.

**Table 14.11-12. Command Line Parameter**

| Parameter Name        | Required | Description                                                                                                                                                                                                                                                                                              |
|-----------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| verifyOnly            | No       | Optional parameter to specify whether to only verify existing checksum. When the option is present in the command line, DPCV will only verify checksum if it is present in the database; When the option is not present, DPCV will calculate a checksum for files that do not have checksum in database. |
| esdts                 | No       | Optional parameter to specify ESDTs needs to be verified. Its value could be keyword ALL (meaning all ESDTs) or a specific list of ESDTS separated by " ". It can not be combined with the file option.                                                                                                  |
| insertBeginTime       | No       | Optional parameter to specify lower limit of granule ingest time used to qualify granules to be verified. It can not be combined with the file option.                                                                                                                                                   |
| insertEndTime         | No       | Optional parameter to specify upper limit of granule ingest time used to qualify granules to be verified. It can not be combined with the file option.                                                                                                                                                   |
| daysSinceLastChecksum | No       | Parameter to specify the cut off value of number of days since the file is last checksummed. Files that are checksummed within the cut off value of days will not be checksummed again.                                                                                                                  |
| file                  | No       | Parameter to specify a list of DataPool granule ids to be verified. It can not be combined with the esdts, insertBeginTime or insertEndTime option.                                                                                                                                                      |
| percentage            | No       | Parameter to specify the percentage of files in the qualifying range that is verified.                                                                                                                                                                                                                   |
| fg                    | No       | Parameter to specify the DPCV process to run as a foreground process. If present, it has to be the first parameter in the parameter list. By default, DPCV will run as a background process. This is reserved for cron job run.                                                                          |
| noprompt              | No       | Parameter to specify the log file name not to be prompted back on the standard out. This is reserved for cron job run.                                                                                                                                                                                   |

### DataPool Checksum Verification Utility Configuration File

The DataPool Checksum Verification utility uses a configuration file: EcDIDpcv.properties, located in /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table 14.11-13 describes the configuration parameters.

**Table 14.11-13. Configuration Parameters**

| Parameter Name              | Description                                                                                                       |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------|
| PGM_ID                      | Program ID used for connecting to the Data Pool database.                                                         |
| SYBASE_USER                 | The user name for the Sybase connection.                                                                          |
| SYBASE_HOST                 | The name of the host Sybase SQL server is on.                                                                     |
| SYBASE_PORT                 | The port number of Sybase server                                                                                  |
| SYBASE_DBNAME               | The name of the Data Pool database you intend to connect to                                                       |
| SYBASE_JDBC_DRIVER_CLASS    | The Sybase jdbc driver class.                                                                                     |
| DB_RETRIES                  | The number of times the utility attempts to connect to the database before exiting. The recommended default is 5. |
| DB_SLEEPSECONDS             | The number of seconds the utility waits ('sleep') between connection attempts. The recommended default is 10.     |
| SQL_TIMEOUT_SECONDS         | The number of seconds to timeout a db operation.                                                                  |
| DPCV_EXPIRATION_TIME        | The number of hours the utility uses to mark an un-finished process as expired.                                   |
| DPCV_HISTORY_RETENTION_TIME | The number of days the utility uses to cleanup old DPCV run record in database.                                   |
| SECONDS_BETWEEN_CHECKSUMS   | The number of seconds between checksum operations                                                                 |
| NUM_CHECKSUM_RETRIES        | The number of retries on checksum failures                                                                        |
| HOST_NAME                   | The host name where DPCV is running                                                                               |

### 14.11.15.1 Running the Data Pool Checksum Verification Utility

- 1 Log in at the machine where the Data Pool Checksum Verification Utility is installed (e.g., x4dpl01).
  - The operator who is executing the script must have privileges for removing science, metadata, and browse files from the Data Pool disks.
- 2 To change to the directory for starting the Data Pool Checksum Verification Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.
- 3 At the prompt, type the command to start the Data Pool Checksum Verification Utility, in the form of the following.
 

```
EcDIDpvcStart [-fg] <MODE> [-verifyOnly] [-esdts (keyword ALL or list of ShortName.VersionId e.g. ALL or "AE_Land.086|PH.001|QA.001")] [-insertBeginTime (MM/DD/YYYY HH:MM:SS)] [-insertEndTime (MM/DD/YYYY HH:MM:SS)] [-daysSinceLastChecksum (number of days)] [-file (text file containing DataPool GranuleIds)] [-percentage (integer from 0-100)] [-noprompt]
```

Data Pool Checksum Verification Utility usage examples:

- For all granules ingested within a period of time run:

```
EcDlDpcvStart OPS -verifyOnly -esdts ALL -insertBeginTime "11/27/2008
00:00:00" -insertEndTime "12/25/2008 23:59:59"
```

The DataPool Checksum Verification utility will perform checksum verification for all granule files ingested between Thanksgiving and Christmas that have existing checksum information.

```
EcDlDpcvStart OPS -verifyOnly -esdts ALL -insertBeginTime "11/27/2008
00:00:00" -insertEndTime "12/25/2008 23:59:59" -percentage 50
```

The DataPool Checksum Verification utility will perform checksum verification for 50% of the granule files ingested between Thanksgiving and Christmas that have existing checksum information.

```
EcDlDpcvStart OPS -verifyOnly -esdts ALL -insertBeginTime "11/27/2008
00:00:00" -insertEndTime "12/25/2008 23:59:59" -daysSinceLastChecksum
30
```

The DataPool Checksum Verification utility will perform checksum verification for all the granule files ingested between Thanksgiving and Christmas that have existing checksum information and haven't been verified for at last 30 days.

- For granules belong to a list of specified ESDTs ingested within a period of time run:

```
EcDlDpcvStart OPS -esdts "AST_L1A.003|MOD29P1D.005" -insertBeginTime
"11/27/2008 00:00:00" -insertEndTime "12/25/2008 23:59:59"
```

The DataPool Checksum Verification utility will perform checksum verification for all granule files that are of ESDT AST\_L1A.003 or MOD29P1D.005 ingested between Thanksgiving and Christmas. If there is no existing checksum information, DPCV will calculate one based on the default checksum type and insert it into the database.

- For a "file" run:

```
EcDlDpcvStart OPS -file dplgranuleids.txt
```

The DataPool Checksum Verification utility will perform checksum verification for all granule files that are listed in the dplgranuleids.txt.

- For a cron run:

```
EcDlDpcvStart -fg OPS -noprompt -verifyOnly -esdts ALL -
daysSinceLastChecksum 30
```

Put the above in the crontab to set up the cron job to verify checksum for files that have not been verified for at least 30 days.

---

### 14.11.16 Running the Restore Online Archive from Tape Utility

The *RestoreOlaFromTape* utility will repair individual granules or files that are lost or damaged in the on-line archive provided that the inventory entries of the corresponding granules are completely intact. This is because *RestoreOlaFromTape* does not have the capability to repair Data Pool inventory database entries. In all other cases, granules must be restored using the Publish utility (e.g., if file entries or browse cross references are missing, or Data Warehouse entries for public granules were damaged or lost). The publish utility has the capability to reconstruct the Data Pool inventory entries for a granule.

The *RestoreOlaFromTape* utility shall:

- Restore defective granules from their tape archive location.
- Verify the checksum of the tape copy.
- Rename the files according to Data Pool rules.
- Restore granule metadata files from the XML file archive.
- Restore browse symbolic links for the science granule that are restored.
- Restore browse granules or files from the browse file archive, which is a disk archive. If the corrupted or lost browse files belong to a public browse granule, the corresponding browse images will be extracted from the original browse file.

In addition, the *RestoreOlaFromTape* utility shall:

- Optimize the restore of the files from the tape archive by organizing them by tape. Within a collection of files from the same tape, files will be organized in ascending block number order. This organization will optimize the tape read operations.
- Allow configurable parallelization of the tape restore operations by providing a configuration parameter that specifies the number of tape drives to be used for the restore operation. Please note that for a given tape, no concurrent/parallel access will be provided. The proposed behavior is based on tape access tests that were conducted for DPAD tape access optimization is ECS release 7.20.
- Manage the capacity demand of bulk repairs to avoid serious degradation of operational workloads (e.g., limits on concurrent tape mounts, tape reads, on-line archive writes, checksumming operations).

Input is provided via an input file.

Table 14.11-14 provides a description of Command Line Parameters.

**Table 14.11-14. Command Line Parameters**

| Parameter Name                              | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>-file &lt;file name&gt;</i>              | Name and path of the input file to be used by the utility                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <i>-contents &lt;contents type&gt;</i>      | <p>The type of contents present in the file. Any of the following options are allowed:</p> <ul style="list-style-type: none"> <li>⇒ dplids: the input file contains the DPL granule IDs or browse IDs of the on-line archive granules that must be repaired</li> <li>⇒ ecsids: the input file contains the ECS granule IDs (dbIDs) or browse IDs of the on-line archive granules that must be repaired</li> <li>⇒ dplfiles: the input file contains the DPL filenames of the files that must be repaired. Browse files in JPG or HDF format are also accepted</li> </ul>                                  |
| <i>[-restoremisbr]</i>                      | <p>Indicates if the utility should restore MISBR browse granule in the DPL. If the parameter is not set, the MISBR browse granule will not be restored.</p> <p>NOTE: This parameter will cause the utility to MISBR browse granule only when the configuration parameter MISR_SPECIAL_PROCESSING is set to "Y".</p>                                                                                                                                                                                                                                                                                       |
| <i>[-restorelinks   -restorelinksonly ]</i> | <p>Indicates if the utility should restore browseQA/PH symbolic linkage file for the given science granule.</p> <p>When '-restorelinks' is not provided in command line, only science granule metadata and data files are restored.</p> <p>When '-restorelinks' is specified in command line, both science granule files and browse/QA/PH symbolic links are restored.</p> <p>When '-restorelinks only' is specified in command line, only browse/QA/PH symbolic links are restored.</p> <p>Note: A list of science granule DPL Ids or ECS Ids should be used to restore browse/QA/PH symbolic links.</p> |
| <i>[-restorexmlonly]</i>                    | <p>Indicates if the utility should restore granule metadata file only.</p> <p>When '-restorexmlonly' is not provided in command line, both granule metadata and data files are restored.</p> <p>When '-restorexmlonly' is specified in command line, only granule metadata file are restored.</p>                                                                                                                                                                                                                                                                                                         |
| <i>[-recovery no]</i>                       | <p>Indicates if the utility should not recover from the last unsuccessful run. By default, the utility will disregard the current input file and read and complete the latest unsuccessful run (request) from the database.</p> <p>NOTE: if NO recovery is desired, the last unsuccessful run will be set to "Aborted" in the database.</p>                                                                                                                                                                                                                                                               |
| <i>[-email recipient_email_address]</i>     | <p>Indicates the Email address of the user to receive the termination status of the utility. Multiple email addresses may be entered, separated by semicolons. If errors occurred, detail about the errors or how to retrieve the details will be present in the Email message.</p>                                                                                                                                                                                                                                                                                                                       |

## Configuration File Format – RestoreOlaFromTape.properties

The configuration file contains vital details about how the utility will operate. The utility will exit immediately if a configuration file is not available. The file is a plain text ASCII file and has the following format. Table 14.11-15 describes the configuration parameters.

**Table 14.11-15. Configuration Parameters (1 of 2)**

| Parameter Name           | Description                                                                                                                                                                                  |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PGM_ID                   | Sybase connectivity, the ID (10000030) is used to decrypt the DB password based on ECS standards                                                                                             |
| SYBASE_HOST              | Sybase connectivity, the host name for the SYBASE data server                                                                                                                                |
| SYBASE_PORT              | Sybase connectivity, the port number for the SYBASE data server on the specified host                                                                                                        |
| SYBASE_USER              | Sybase connectivity, the user name (EcDIRestoreOlaFromTape) used to login to the SYBASE data server. DPL and AIM databases                                                                   |
| SYBASE_AIM_DBNAME        | Sybase connectivity                                                                                                                                                                          |
| SYBASE_DPL_POOL_SIZE     | Sybase connectivity, the database connection pool size for the DPL                                                                                                                           |
| SYBASE_AIM_POOL_SIZE     | Sybase connectivity, the database connection pool size for the AIM                                                                                                                           |
| SYBASE_JDBC_DRIVER_CLASS | Sybase connectivity                                                                                                                                                                          |
| DB_RETRIES               | Number of retries of a RETRYABLE DB operation (e.g. deadlock)                                                                                                                                |
| DB_SLEEPSECONDS          | Number of sleep seconds between retries                                                                                                                                                      |
| SQL_TIMEOUT_SECONDS      | Time in seconds that an SQL query will execute before timing out.                                                                                                                            |
| DEBUG_MESSAGES           | (Y/N) indicates if detailed debugging information will be written to the log file.                                                                                                           |
| CHECKSUM_SERVICE_HOSTS   | The service hosts to be used for checksumming. The service hosts are configured in the format of <host_name_1>:<port_num>:<num_of_slots_1>, <host_name_2>:<port_num>:<num_of_slots_2>, ...   |
| CHECKSUM_TIMEOUT         | Number of seconds before timeout a checksum operation                                                                                                                                        |
| COPY_SERVICE_HOSTS       | The service hosts to be used for copy operation. The service hosts are configured in the format of <host_name_1>:<port_num>:<num_of_slots_1>, <host_name_2>:<port_num>:<num_of_slots_2>, ... |
| COPY_TIMEOUT             | Number of seconds before timeout a copy operation                                                                                                                                            |
| SNSM_QS_HOST             | StorNext Metadata Server Quick Server host                                                                                                                                                   |
| SNSM_QS_PORT             | StorNext Metadata Server Quick Server port                                                                                                                                                   |
| CONNECT_QS_RETRIES       | Number of retries for Quick Server call failures                                                                                                                                             |
| CONNECT_QS_RETRY_SECONDS | Number of sleep seconds between the retries of a Quick Server call                                                                                                                           |

**Table 14.11-15. Configuration Parameters (2 of 2)**

| Parameter Name              | Description                                                                                                                                                                                                                                                                                                                                                                                      |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| COPY_BLOCK_SIZE_KBYTES      | copy block size used by EcAdCopy                                                                                                                                                                                                                                                                                                                                                                 |
| COPY_RETRIES                | number of retries for EcAdCopy on read/write failures                                                                                                                                                                                                                                                                                                                                            |
| REQUEST_RETENTION_DAYS      | The request retention time in days                                                                                                                                                                                                                                                                                                                                                               |
| EMAIL_SMTP_HOST             | The Email SMTP server host                                                                                                                                                                                                                                                                                                                                                                       |
| EMAIL_FROM_ADDRESS          | Outbound email from address to operator                                                                                                                                                                                                                                                                                                                                                          |
| DEDICATED_TAPE_DRIVES       | Number of tape drives (tapes) that can be concurrently used for restores. Recommended values 4 or DAAC defined.                                                                                                                                                                                                                                                                                  |
| CONCURRENT_RESTORES         | Number of restores that can be issued concurrently for a given drive containing a restore tape. The restores will not happen concurrently per say but they will be enqueued by the tape management COTS and will be executed concurrently. The parameter optimizes tape reads by preventing the tape from being stopped during the restore. Recommended values can be anywhere between 5 and 10. |
| DTD_VERSION                 | DTD Version of xml files for DAP, PH, QA granules                                                                                                                                                                                                                                                                                                                                                |
| DATA_CENTER_ID              | DATA_CENTER_ID of xml files for DAP, PH, QA granules                                                                                                                                                                                                                                                                                                                                             |
| CONCURRENT_GET_FILETAPEINFO | Number of threads that can be issued concurrently when retrieving and updating file tape information. Recommended values 10.                                                                                                                                                                                                                                                                     |
| MISR_SPECIAL_PROCESSING     | controls if MISR Browse special processing module is ON (Y) or OFF (N)                                                                                                                                                                                                                                                                                                                           |

### 14.11.16.1 Running the Restore Online Archive from Tape Utility

- 1 Log in at the machine where the Restore Online Archive from Tape Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Restore Online Archive from Tape Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the Restore Online Archive from Tape Utility, in the form of the following.

```
EcDlRestoreOlaFromTapeStart -mode <mode> -file <file name and path with contents specified by -contents parameter> -contents dplids | ecsids | dplfiles [-restoremisbr] [-restorelinks | -restorelinksonly] [-restorexmlonly] [-recovery no] [-email <usertoreceivestatusemail>]
```

Restore Online Archive from Tape Utility usage examples:

- **EcDlRestoreOlaFromTapeStart** -mode OPS -file </home/john/dplids.txt> -contents dplids -recovery no -email cmshared@ecs.nasa.gov

Restores to the on-line archive from tape the DPL granules with the DPL IDs present in the dplids.txt flat file. The utility will NOT recover from an unsuccessful previous

run and will set the restore from tape request to “Aborted” in the DPL database for the unsuccessful previous run. An Email with the request status will be sent to the `cmshared@ecs.nasa.gov` once the utility finishes the current request.

- `EcDlRestoreOlaFromTapeStart -mode OPS -file </home/john/ecsids.txt> -contents ecsids -recovery no`

Restores to the on-line archive from tape the DPL granules with the ECS IDs present in the `ecsids.txt` flat file. The utility will NOT recover from an unsuccessful previous run and will set the restore from tape request to “Aborted” in the DPL database for the unsuccessful previous run.

- `EcDlRestoreOlaFromTapeStart -mode OPS -file </home/john/dplfiles.txt> -contents dplfiles -recovery no`

Restores to the on-line archive from tape the DPL files with the full path and filenames specified in the `dplfiles.txt` flat file. The utility will NOT recover from an unsuccessful previous run and will set the restore from tape request to “Aborted” in the DPL database for the unsuccessful previous run.

- `EcDlRestoreOlaFromTapeStart -mode OPS -file </home/john/dplids.txt> -contents dplids`

Reruns the previous unsuccessful restore from tape request based on the information saved in the DPL database tables used by the utility. The current input file is NOT used. In order to restore the granules specified in the input file, the utility must be restarted after the recovery run completes.

---

#### 14.11.17 Running the Restore Tape from Online Archive Utility

The *RestoreTapeFromOla* utility will repair individual files that are lost or corrupted on tape based on the primary file instance that is present in the on-line archive. The files being restored must be inventoried both in the AIM and DPL databases because the utility does not create new AIM or DPL database entries. The utility shall:

- Allow DAAC staff to replace individual granules in the tape archive from their on-line copy (after verification that the on-line copy is still intact). Files will be renamed appropriately to conform to the tape archive naming conventions.
- Manage the capacity demand of bulk repairs to avoid serious degradation of operational workloads (e.g., limits on concurrent tape mounts, on-line archive reads, tape writes, and checksumming operations).

Notes:

- Since the on-line Browse archive is not part of the Data Pool, this repair function will not cover Browse archive repairs. They can be repaired using StorNext utilities like today.

- The **RestoreTapeFromOla** utility will not cover XML metadata files. The XML file archive restore function is performed using other procedures.

Input is provided via an input file. Table 14.11-16 provides a description of Command Line Parameters.

**Table 14.11-16. Command Line Parameters (1 of 2)**

| Parameter Name            | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| -file <file name>         | Name and path of the input file to be used by the utility                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| -contents <contents type> | <p>The type of contents present in the file. Any of the following options are allowed:</p> <ul style="list-style-type: none"> <li>⇒ mediaids: the input file contains the media IDs (tape labels) of the tapes that were lost / damaged.</li> <li>⇒ tapefiles: the input file contains the complete file names and paths of the tape files that must be repaired.</li> <li>⇒ dplids: the input file contains the DPL granule IDs of the tape granules that must be repaired</li> <li>⇒ ecsids: the input file contains the ECS granule IDs (dbIDs) of the tape granules that must be repaired</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| [-removereadonlyfile]     | <p>Indicates that the utility should remove the original tape file from archive if the file cannot be restored to its original location. The utility always restores the file to the currently opened volume groups. Details below:</p> <p>If the option is not present, the utility will not try to remove the original tape file from archive. If the file cannot be restored to its original location, it will be restored in the currently opened volume group with the new file name, and the original file will remain on tape at the original location, without any corresponding AIM inventory record. The utility will not even try to remove the original file, regardless of the permissions on it.</p> <p>If the option is present, the utility will try to remove the original file. The file restored has the same name as original file. The utility will prompt the user to verify that the permissions to the RO volume group have been changed to RW if necessary:</p> <p>Have you changed the RO permissions to RW in the RO volume group affected by the restore (Y/N)?</p> <p>On Y the utility will proceed and:</p> <p>If the permissions to the affected files are RW, it will remove the original files that are affected.</p> <p>If the permissions to the affected files are RO, it will FAIL the restore of the granules involved. It is the responsibility of DAAC operations to inspect the log, identify the failed granules and rerun the utility after setting the correct RW permissions to the closed Volume Group. The reason for the failure is that if we would in fact restore the granule, the original file will remain on tape at the original location, and other application will find the bad copy.</p> <p>On N the utility will exit.</p> |

**Table 14.11-16. Command Line Parameters (2 of 2)**

| Parameter Name                          | Description                                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>[-recovery no]</i>                   | Indicates that the utility should not recover from the last unsuccessful run. By default, the utility will disregard the current input file and read and complete the latest unsuccessful run (request) from the database.<br><br>NOTE: if NO recovery is desired, the last unsuccessful run will be set to "Aborted" in the database. |
| <i>[-email recipient_email_address]</i> | Indicates the Email address of the user to receive the termination status of the utility. Multiple email addresses may be specified, separated by semicolons. If errors occurred, detail about the errors or how to retrieve the details will be present in the Email message.                                                         |

### Configuration File Format – RestoreTapeFromOla.properties

The configuration file contains vital details about how the utility will operate. The utility will exit immediately if a configuration file is not available. The file is a plain text ASCII file and has the following format. Table 14.11-17 describes the configuration parameters.

**Table 14.11-17. Configuration Parameters (1 of 2)**

| Parameter Name             | Description                                                                                                                |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------|
| PGM_ID                     | Sybase connectivity, the ID (10000031) is used to decrypt the DB password based on ECS standards                           |
| SYBASE_HOST                | Sybase connectivity, the host name for the SYBASE data server                                                              |
| SYBASE_PORT                | Sybase connectivity, the port number for the SYBASE data server on the specified host                                      |
| SYBASE_USER                | Sybase connectivity, the user name (EcDIRestoreTapeFromOla) used to login to the SYBASE data server. DPL and AIM databases |
| SYBASE_DPL_DBNAME          | Sybase connectivity, the database name for the DPL database                                                                |
| SYBASE_AIM_DBNAME          | Sybase connectivity AIM database name                                                                                      |
| SYBASE_DPL_POOL_SIZE       | Sybase connectivity, the database connection pool size for the DPL                                                         |
| SYBASE_AIM_POOL_SIZE       | Sybase connectivity, the database connection pool size for the AIM                                                         |
| SYBASE_JDBC_DRIVER_CLASSES | Sybase connectivity, JDBC driver class                                                                                     |
| DB_RETRIES                 | Number of retries of a RETRYABLE DB operation (e.g. deadlock)                                                              |
| DB_SLEEPSECONDS            | Number of sleep seconds between retries                                                                                    |
| SQL_TIMEOUT_SECONDS        | Time in seconds that an SQL query will execute before timing out.                                                          |
| DB_BATCH_SIZE              | The batch size for the database retrieve operations, its default value is 50                                               |

**Table 14.11-17. Configuration Parameters (2 of 2)**

| Parameter Name                       | Description                                                                                                                                                                                                                                                                                                       |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DEBUG_MESSAGES                       | (Y/N) indicates if detailed debugging information will be written to the log file.                                                                                                                                                                                                                                |
| CHECKSUM_SERVICE_HOSTS               | The service hosts to be used for checksumming.<br>The service hosts are configured in the format of <host_name_1>:<port_num>:<num_of_slots_1>, <host_name_2>:<port_num>:<num_of_slots_2>, ...                                                                                                                     |
| CHECKSUM_TIMEOUT                     | Number of seconds before timeout a checksum operation                                                                                                                                                                                                                                                             |
| COPY_SERVICE_HOSTS                   | The service hosts to be used for copy operation.<br>The service hosts are configured in the format of <host_name_1>:<port_num>:<num_of_slots_1>, <host_name_2>:<port_num>:<num_of_slots_2>, ...                                                                                                                   |
| COPY_TIMEOUT                         | Number of seconds before timeout a copy operation                                                                                                                                                                                                                                                                 |
| SNSM_QS_HOST                         | StorNext Metadata Server Quick Server host                                                                                                                                                                                                                                                                        |
| SNSM_QS_PORT                         | StorNext Metadata Server Quick Server port                                                                                                                                                                                                                                                                        |
| SNSM_QS_OUTPUT_DIR                   | The directory where StorNext Metadata Server Quick Server puts the output files. The directory should be visible from both the host where the StorNext Metadata Server Quick Server runs and from the host where the RestoreTapeFromOla utility runs. The directory should not be shared with other applications. |
| CONNECT_QS_RETRIES                   | Number of retries for Quick Server call failures                                                                                                                                                                                                                                                                  |
| CONNECT_QS_RETRY_SECONDS             | Number of sleep seconds between the retries of a Quick Server call                                                                                                                                                                                                                                                |
| COPY_BLOCK_SIZE_KBYTES               | copy block size used by the copy utility                                                                                                                                                                                                                                                                          |
| COPY_RETRIES                         | number of retries for the copy utility on read/write failures                                                                                                                                                                                                                                                     |
| REQUEST_RETENTION_DAYS               | The request retention time in days                                                                                                                                                                                                                                                                                |
| EMAIL_SMTP_HOST                      | The Email SMTP server host                                                                                                                                                                                                                                                                                        |
| EMAIL_FROM_ADDRESS                   | Outbound email from address to operator                                                                                                                                                                                                                                                                           |
| CONCURRENT_TAPE_ARCHIVE_CACHE_WRITES | Number of concurrent writes to the tape archive cache. This is a throttling mechanism that controls how many files can be concurrently copied from the on-line archive to tape.                                                                                                                                   |

### 14.11.17.1 Running the Restore Tape from Online Archive Utility

- 1 Log in at the machine where the Restore Tape from Online Archive Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Restore Tape from Online Archive Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the Restore Tape from Online Archive Utility, in the form of the following.

```
EcDlRestoreTapeFromOlaStart -mode <mode> -file <file name and path of input file whose contents type is specified by the -contents parameter> -contents <mediuids | tapefiles | dplids | ecsids> [-recovery no] [-email <usertoreceivestatusemail>]
```

Restore Tape from Online Archive Utility usage examples:

- `EcDlRestoreTapeFromOlaStart -mode OPS -file  
</home/john/mediads.txt> -contents mediaids -recovery no -email  
cmshared@ecs.nasa.gov`

Restores all files on the tape(s) specified in the mediaids.txt input file from their on-line archive copy. The utility will NOT recover from an unsuccessful previous run and will set the previous restore on-line archive to tape request to “Aborted” in the DPL database. An Email with the request status will be sent to the cmshared@ecs.nasa.gov once the utility finishes the current request.

- `EcDlRestoreTapeFromOlaStart -mode OPS -file  
</home/john/tapefiles.txt> -contents tapefiles -recovery no`

Restores the tapes files specified in the tapefiles.txt input file from their on-line archive copy. The utility will NOT recover from an unsuccessful previous run and will set the previous restore on-line archive to tape request to “Aborted” in the DPL database.

- `EcDlRestoreTapeFromOlaStart -mode OPS -file  
</home/john/dplids.txt> -contents dplids -recovery no`

Restores the granules with the DPL IDs specified in the dplids.txt input file from their on-line archive copy. The utility will NOT recover from an unsuccessful previous run and will set the previous restore on-line archive to tape request to “Aborted” in the DPL database.

- `EcDlRestoreTapeFromOlaStart -mode OPS -file  
</home/john/ecslids.txt> -contents ecsids`

Restores the granules with the ECS IDs specified in the ecsids.txt input file from their on-line archive copy. If there was an unsuccessful previous run, the utility will recover from that run based on the information saved in the DPL database tables used by the utility, and the current input file will not be used. The current runs must be restarted after the recovery run is completed.

---

### 14.11.18 Running the Archive Checksum Verification Utility

The Archive Checksum Validation utility (ACVU) provides a mechanism by which the ECS Operations Staff can perform checksum verification of files in the AIM archive. The utility allows the operator to specify which files to verify, by sampling files based on media ID (a single media ID or a list of media IDs), volume group (a single volume group or a list of volume groups), or granule ID (a single granule ID, a list of granule IDs, or an input file containing granule IDs). The operator may also restrict verification to files which have not had their checksum verified within an operator-specified time period.

According to the sampling criteria specified, the utility will identify the files to be verified, organize the result by location on tape, verify their checksum values, and update the last

checksum verification time and status in the AIM Inventory database. The utility will need to verify that an LTO tape is in the near-line archive (i.e. not off-line) and alert the operator if the tape is off-line.

Upon detection of checksum verification failure after a configurable number of retry attempts (NUM\_CHECKSUM\_RETRIES in configuration file), the utility will log detailed information about the failure including media ID, volume group, granule ID, ESDT, insert time, complete file path and file name, along with the checksum information -- including checksum type, checksum values (computed value vs the corresponding value stored in database), the last time the file was checksummed, and checksum origin (who performed the last checksum). The checksum status of the file will be updated in the AIM Inventory database to mark it as a case of checksum verification failure.

The log will also include statistical summary information including total number of files checked, number of files that failed checksum, percentage of files that failed checksum, categorized by ESDT. This utility is designed such that the checksum verification can be throttled (by adjusting the number of concurrent tapes and number of concurrent tape reads) so it does not impact on-going daily operations. Table 14.11-18 provides a description of Command Line Parameters.

**Table 14.11-18. Command Line Parameter (1 of 2)**

| <b>Parameter Name</b> | <b>Required</b>                                                     | <b>Description</b>                                                                                                                                                                                              |
|-----------------------|---------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| calculate             | No                                                                  | Optional parameter to specify whether to calculate and store checksums for files found currently without checksums.                                                                                             |
| days                  | No                                                                  | Optional parameter to specify days since last checked.                                                                                                                                                          |
| percent               | No                                                                  | Optional parameter to specify percentage of files to check.                                                                                                                                                     |
| norecovery            | No                                                                  | Optional parameter to specify not to recover from previous run.                                                                                                                                                 |
| volume group          | Yes, if mediaid, granuleid, or file parameters are not present      | Parameter to specify volume groups whose files will have their checksum verified. This is a comma separated list of one or more volume groups (no spaces). Volume groups should be specified by full path name. |
| mediaid               | Yes, if volume group, granuleid, or file parameters are not present | Parameter to specify mediaids whose files will have their checksum verified. This is a comma separated list of one or more mediaids (no spaces).                                                                |
| granuleid             | Yes, if volume group, mediaid, or file parameters are not present   | Parameter to specify granules whose files will have their checksum verified. This is a comma separated list of one or more granule ids (no spaces).                                                             |

**Table 14.11-18. Command Line Parameter (2 of 2)**

| Parameter Name | Required                                                              | Description                                                                                                                                                                                          |
|----------------|-----------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| file           | Yes, if volumegroup, mediaid, or granuleid parameters are not present | Parameter to specify the name of an input file containing granuleids of granules whose files will have their checksum verified. Granuleids should be listed in the input file separated by newlines. |
| outputDir      | No                                                                    | Parameter to specify directory for error files under /workingdata/emd/<MODE>/Acvu                                                                                                                    |

### Archive Checksum Validation Utility Configuration File

The Archive Checksum Validation utility uses a configuration file, EcDsAmAcvu.CFG, located in /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table 14.11-19 describes the configuration parameters.

**Table 14.11-19. Configuration Parameters (1 of 2)**

| Parameter Name          | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SYB_USER                | Sybase login name for the user of the Inventory database.                                                                                                                                                                                                                                                                                                                                                                                                          |
| SYB_SQL_SERVER          | Name of Sybase SQL Server hosting Inventory database.                                                                                                                                                                                                                                                                                                                                                                                                              |
| SYB_DBNAME              | Name of Inventory database.                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| PGM_ID                  | Program identifier used as seed to generate database password.                                                                                                                                                                                                                                                                                                                                                                                                     |
| NUM_RETRIES             | Number of times database operation will be attempted.                                                                                                                                                                                                                                                                                                                                                                                                              |
| RETRY_INTERVAL          | Number of seconds between retries.                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| SNSM_HOST               | The Stornext host                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| SNSM_PORT               | The Stornext port                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| SNSM_TEMP_DIR           | The directory to place file listings for tapes. This directory should be cross mounted between the Stornext host and the oml host. The suggested directory is /workingdata/emd/<MODE>/Acvu/TempDir. The directory should be readable by cmshared with write permissions for the Stornext user (smuser). To achieve this we suggest having the directory owned by smuser, a groupid of cmshared, and 775 permissions. This directory should be cleaned up manually. |
| MAX_BLOCKINFO_PROCESSES | Number of processes to get block info from media concurrently.                                                                                                                                                                                                                                                                                                                                                                                                     |
| MAX_TAPE_READS          | Number of read requests per tape at once                                                                                                                                                                                                                                                                                                                                                                                                                           |
| MAX_CONCUR_TAPES        | Number of tapes that can be read from at once                                                                                                                                                                                                                                                                                                                                                                                                                      |

**Table 14.11-19. Configuration Parameters (2 of 2)**

| Parameter Name        | Description                                                                                                                                                        |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NUM_CHECKSUM_RETRIES  | Number of times a checksum will be attempted.                                                                                                                      |
| VALIDATION_OUTPUT_DIR | The default directory to place error output files. The directory should be readable/writable by cmshared. The suggested directory is /workingdata/emd/<MODE>/Acvu. |

### 14.11.18.1 Running the Archive Checksum Verification Utility

Log in at the machine where the Archive Checksum Verification Utility is installed (e.g., x4oml01).

- 2 To change to the directory for starting the Archive Checksum Verification Utility, type `cd /usr/ecs/<MODE>/CUSTOM/utilities` and then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the Archive Checksum Verification Utility, in the form of the following.

```
EcDsAmAcvu.pl <MODE> [-calculate]

 [-days <NUMBER OF DAYS>]

 [-percent <PERCENT 1-100>]

 [-norecovery]

 (-volumegroup <VOLUME GROUPS> |
 -mediaid <MEDIAIDS> |
 -granuleid <GRANULEIDS> |
 -file <FILENAME>) |
 [-outputDir <DIRECTORY>]
```

Archive Checksum Verification Utility usage examples:

- For a "volumegroup" run:

```
EcDsAmAcvu.pl OPS -volumegroup /stornext/snfs1/OPS/MODIS
```

The Archive Checksum Validation Utility will perform checksumming for all files in specified volumegroup (/stornext/snfs1/OPS/MODIS).

```
EcDsAmAcvu.pl OPS -volumegroup

 /stornext/snfs1/OPS/MODIS,/stornext/snfs1/OPS/ASTER -percent 50
```

The Archive Checksum Validation Utility will perform checksumming for 50% of the files in the specified volume groups.

- For a "media id" run:

```
EcDsAmAcvu.pl OPS -mediaid VG7029
```

The Archive Checksum Validation Utility will perform checksumming for all files on the specified tape

```
EcDsAmAcvu.pl OPS -mediaid VG7029,TG8024 -days 10
```

The Archive Checksum Validation Utility will perform checksumming for the files on the specified tapes which have not been verified in the past 10 days.

- For a "granuleid" run:

```
EcDsAmAcvu.pl OPS -granuleid 22083,22085,22087
```

The Archive Checksum Validation Utility will perform checksumming for the files related to granules 22083, 22085, and 22087 in OPS mode

```
EcDsAmAcvu.pl OPS -granuleid 22083,22085,22087 -calculate
```

The Archive Checksum Validation Utility will perform checksumming for the files related to granules 22083, 22085, and 22087 in OPS mode and if the files do not have a checksum, one will be calculated for it.

- For a "file" run:

```
EcDsAmAcvu.pl OPS -file granuleids.txt
```

The Archive Checksum Validation Utility will perform checksumming for the files related to granules specified in granuleids.txt

```
EcDsAmAcvu.pl OPS -file granuleids.txt -norecovery
```

The Archive Checksum Validation Utility will ignore recovery for any previous run and perform checksumming for the files related to granules specified in granuleids.txt

---

#### **14.11.19 Running the XML Check Utility**

The XML Check utility provides a mechanism by which the ECS Operations Staff can periodically check for corruption in the XML Archive.

In order to detect corruption, the utility verifies the contents of the files are well formed using xmllint.

There are seven command line parameters that may be used. Table 14.11-20 provides a description of Command Line Parameters.

**Table 14.11-20. Command Line Parameter**

| Parameter Name | Required                                             | Description                                                                                                                         |
|----------------|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| days           | No                                                   | Optional parameter to specify days since last checked.                                                                              |
| percent        | No                                                   | Optional parameter to specify percentage of files to check                                                                          |
| ESDT           | Yes, if granuleid or file parameters are not present | Parameter to specify which ESDTs to check. This is a comma separated list (no spaces). Can also specify "ALL" to include all ESDTs. |
| startdate      | No                                                   | Optional parameter used with –ESDT option. Specifies starting insert date to use for ESDTs.                                         |
| enddate        | No                                                   | Optional parameter used with –ESDT option. Specifies ending insert date to use for ESDTs.                                           |
| granuleid      | Yes, if ESDT or file parameters are not present      | Parameter to specify which granules to check. This is a comma separated list (no spaces).                                           |
| file           | Yes, if ESDT or granuleid parameters are not present | Parameter to specify which granules to check. Granule ids should be listing in a file separated by newlines.                        |
| outputDir      | No                                                   | Parameter to specify directory for error files under /workingdata/emd/<MODE>/Xcu                                                    |

### XML Check Configuration File

The XML Check utility uses a configuration file, EcDsAmXcu.CFG, located in /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table 14.11-21 describes the configuration parameters.

**Table 14.11-21. Configuration Parameters (1 of 2)**

| Parameter Name | Description                                                    |
|----------------|----------------------------------------------------------------|
| SYB_USER       | Sybase login name for the user of the Inventory database.      |
| SYB_SQL_SERVER | Name of Sybase SQL Server hosting Inventory database.          |
| SYB_DBNAME     | Name of Inventory database.                                    |
| PGM_ID         | Program identifier used as seed to generate database password. |
| NUM_RETRIES    | Number of times database operation will be attempted.          |
| RETRY_INTERVAL | Number of seconds between retries.                             |

**Table 14.11-21. Configuration Parameters (2 of 2)**

| Parameter Name        | Description                                                                                                                                                      |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MAX_CONCUR_CHECKS     | Number of concurrent calls to xmllint that will be allowed.                                                                                                      |
| VALIDATION_OUTPUT_DIR | The default directory to place error output files. The directory should be readable/writable by cmshared. The suggested directory is /workingdata/emd/<MODE>/Xcu |

### 14.11.19.1 Running the XML Check Utility

- 1 Log in at the machine where the XML Check Utility is installed (e.g., x4oml01).
- 2 To change to the directory for starting the XML Check Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the XML Check Utility, in the form of the following.  
The XML Check utility should be started by the user cmshared (or similar). The XML Check utility is started by entering the following command:

**EcDsAmXcu.pl <mode> <command line parameters>**

XML Check Utility usage examples:

- For an "ESDT" run:

```
EcDsAmXcu.pl OPS -ESDT ALL
```

The XML Check Utility will perform checking for all xml files in OPS mode

```
EcDsAmXcu.pl OPS -ESDT AST_L1A.003,MOD29.005 -startdate Jan 20 2008
-enddate Dec 1 2008
```

The XML Check Utility will performed checking for all AST\_L1A.003 and MOD29.005 xml files whose granules have been inserted between Jan 20 2008 and December 1 2008.

```
EcDsAmXcu.pl OPS -ESDT AST_L1B.003 -percent 50 -days 10
```

The XML Check Utility will perform checking for 50% of AST\_L1B.003 granules which have not been checked in the last 10 days.

- For a "granuleid" run:

```
EcDsAmXcu.pl OPS -granuleid 22083,22085,22087
```

The XML Check Utility will perform checking for the xml files related to granules 22083, 22085, and 22087in OPS mode

- For a "file" run:

```
EcDsAmXcu.pl OPS -file granuleids.txt
```

The XML Check Utility will perform checking for the xml files related to granules specified in granuleids.txt

---

#### 14.11.20 Running the Data Pool Band Backfill Utility

The DPL Backfill Utility is a command line tool that can correct band extraction problems that occurred during DPL registrations. Granule registrations cannot fail if band extraction problems are encountered but the subsequent publications on convert-enabled data types must fail if the band information is not present in the Inventory database at publication time.

The Band Backfill utility was developed to correct the problems above. It will:

- backfill the band information in the Inventory database for the registered granules specified in its input file.
- request the publication of the backfilled granules via the new Data Pool Action driver.

The DAAC Operations staff can identify the granules that need band backfill via the Data Pool Maintenance GUI or by inspecting the EcDINewInsertUtilityDPAD.log file. In both cases, the type of error encountered is:

*ERROR pubreg operation encountered a convertEnabled granule with no band information, granuleState*

For each Data Pool granuleId in its input file, the utility will perform the following steps:

1. Validate that the granule is in the hidden Data Pool. The granules can belong to DPL Ingest (IsOrderOnly = H) or to OMS (IsOrderOnly = Y).
2. Validate that the granule belongs to a convert-enabled ESDT.
3. Validate that the Inventory database contains no band information for this granule.
4. Extract the band information from the granule data files and produce a .BandHeader file. This step is performed by invoking an external script (./custom/utilities/EcDIAdHEGStart). The same script is also used by the new Data Pool Action Driver to create the .BandHeader file during granule registrations. Note: for a multi-file granule, the first file that contains band information will be used.
5. Parse the .BandHeader file and insert the necessary information in the Data Pool database. The .BandHeader file will be removed once it has been parsed.
6. Request the publication of the backfilled granule by inserting a record in the DLIInsertActionQueue table in the Inventory database.

7. Process the next granule in the input file. Note: if an error is encountered during the processing of a granule, the error is logged and the utility continues with processing of the subsequent granules.

Table 14.11-22 provides a description of Command Line Parameters.

**Table 14.11-22. Command Line Parameters**

| Parameter Name     | Description                                                                                                                                                                                                                             |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| -mode <mode>       | Specifies the mode of operation (OPS, TS1, etc.)                                                                                                                                                                                        |
| -file <input file> | Specifies the full path and file name of the file containing the Data Pool granule IDs of the granules that need to be populated with band information. The file is a flat ASCII file and it contains one Data Pool granuleId per line. |

### 14.11.20.1 Running the Data Pool Band Backfill Utility

---

- 1 Log in at the machine where the Data Pool Band Backfill Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Data Pool Band Backfill Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the Data Pool Band Backfill Utility, in the form of the following:

- **EcDlBandBackfillUtilityStart -mode <mode> -file <input file>**

Data Pool Band Backfill Utility usage examples:

- **EcBandBackfillUtilityStart -mode OPS -file /home/cmshared/granuleIds.txt**

Backfills the band information and requests the DPL publication for the granuleIds contained in the specified file. The file contains one Data Pool granuleId per line.

---

### 14.11.21 Running the Data Pool Remove Collection Utility

The Data Pool Remove Collection utility provides a mechanism by which ECS Operations staff can remove the Online Archive configuration of collections from the Inventory database that are no longer of interest to the end users. This utility should be followed by removing the ESDT in the AIM ESDT Maintenance GUI. Table 14.11-23 provides a description of Command Line Parameters.

**Table 14.11-23. Command Line Parameters**

| Parameter Name | Required | Description                                                                                                                                                            |
|----------------|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| debug          | No       | Helps developers debug the app by printing copious debug information                                                                                                   |
| ShortName      | Yes      | ShortName of the collection to be deleted                                                                                                                              |
| VersionId      | Yes      | VersionId of the collection being deleted                                                                                                                              |
| inpfiler       | Yes      | The full path to an input file specifying multiple collections. Please note that either an input file or a ShortName/VersionId combination should be used but not both |

### Remove Collection Configuration File

The Data Pool Remove Collection utility uses a configuration file, EcDIRemoveCollection.CFG, located in the /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table 14.11-24 describes the configuration parameters.

**Table 14.11-24. Configuration Parameters**

| Parameter Name | Value Description                                              |
|----------------|----------------------------------------------------------------|
| SYB_USER       | Sybase login name for the user of the Data Pool database.      |
| SYB_SQL_SERVER | Name of Sybase SQL Server hosting Data Pool database.          |
| SYB_DBNAME     | Name of Data Pool database.                                    |
| PGM_ID         | Program identifier used as seed to generate database password. |
| NUM_RETRIES    | Number of times database operation will be attempted.          |
| SLEEP_SEC      | Number of seconds between retries.                             |

#### 14.11.21.1 Running the Data Pool Remove Collection Utility

- 1 Log in at the machine where the Data Pool Remove Collection Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Data Pool Remove Collection Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the Data Pool Remove Collection Utility, in the form of the following (perform only one command):

- **EcDIRemoveCollection.pl <MODE> -ShortName <SHORTNAME> -VersionId <VERSIONID> [-debug]**

OR

- **EcDIRemoveCollection.pl <MODE> -inpfiler <INPUTFILENAME> [-debug]**

Data Pool Remove Collection Utility usage examples:

- **EcDIRemoveCollection.pl OPS -ShortName AST\_L1A -VersionId 003 -debug**

Remove the Online Archive configuration settings from the collection from the Inventory database. Also print copious debug information.

- `EcDlRemoveCollection.pl OPS -inpfiler /home/cmshared/collections.txt -debug`

Remove the Online Archive configuration settings for multiple collections specified in input file. Also print copious debug information

---

### 14.11.22 Running the Data Pool Cloud Cover Utility

The command line script enables the DAAC to correct the cloud cover (CC) information stored in the DPL database for public science granules, based on the cloud cover source that is associated to the desired collection via the DPL Maintenance GUI.

The script also modifies the 7.22 cloud cover derivation for granules that have multiple measured parameters and different associated cloud cover values for those parameters. The new derivation chooses the maximum cloud cover value and associates it to the granules as opposed to the old derivation that used the first cloud cover value in the granule XML file for the association.

DAAC operations will determine when the best time to run it is and what collections they want to run it for.

The Data Pool Cloud Cover Utility script uses a configuration file containing details about how to connect to Sybase, the number of threads that are used concurrently to perform the requested cloud cover operation and the level of verbosity in the log file. The file *EcDlCloudCoverUtility.properties* contains the configuration parameters for the Data Pool Cloud Cover Utility script. The configuration file must be a plain text ASCII file, which has the following format:

```
PGM_ID=10000033
SYBASE_HOST=<string>
SYBASE_PORT=<integer>
SYBASE_DPL_DBNAME=<string>
SYBASE_USER=EcDlCloudCoverUtility
SYBASE_JDBC_DRIVER_CLASS=com.sybase.jdbc3.jdbc.SybDriver
DB_RETRIES=<integer>
DB_SLEEPSECONDS=<integer>
DB_BATCH_SIZE=<integer>
SQL_TIMEOUT_SECONDS=<integer>
concurrent number of threads getting Cloud Cover PSA value from xml file
CONCURRENT_GET_CLOUDCOVERP=<integer>
DEBUG_MESSAGES=<string Y/N>
```

Table 14.11-25 describes the individual configuration parameters mentioned above:

**Table 14.11-25. Data Pool Access Configuration Parameters for Cloud Cover Scripts**

| Parameter Name             | Description                                                                                                                                                                                                                                                                                                                                                                                     |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PGM_ID                     | Used to decrypt the database connection password for the database user represented by this utility.                                                                                                                                                                                                                                                                                             |
| SYBASE_HOST                | The Sybase database host to connect to.                                                                                                                                                                                                                                                                                                                                                         |
| SYBASE_PORT                | The Sybase database port on the specified host.                                                                                                                                                                                                                                                                                                                                                 |
| SYBASE_DPL_DBNAME          | The name of the Data Pool database to connect to.                                                                                                                                                                                                                                                                                                                                               |
| SYBASE_USER                | The user name for the Sybase connection.                                                                                                                                                                                                                                                                                                                                                        |
| SYBASE_JDBC_DRIVER_CLASS   | The name of the SQL server for this Sybase connection.                                                                                                                                                                                                                                                                                                                                          |
| DB_RETRIES                 | The number of times the utility attempts to connect to the database. The recommended default is 5.                                                                                                                                                                                                                                                                                              |
| DB_SLEEPSECONDS            | The number of seconds the utility waits ("sleeps") between connection attempts. The recommended default is 15.                                                                                                                                                                                                                                                                                  |
| DB_BATCH_SIZE              | The batch size for processing the database requests. The recommend default is 2000.                                                                                                                                                                                                                                                                                                             |
| SQL_TIMEOUT_SECONDS        | The number of seconds after which a pending / hung SQL command will be time out. The recommend default is 3600.                                                                                                                                                                                                                                                                                 |
| CONCURRENT_GET_CLOUDCOVERP | The number of concurrent threads used when executing operations (populate or repopulate) of the cloud cover collections for which the cloud cover source is a PSA. Concurrency is needed for performance reasons because the cloud cover value must be read from the granule XML file. Please note that the utility will use CONCURRENT_GET_CLOUDCOVERP + 1 database connections while it runs. |
| DEBUG_MESSAGES             | Y – provides verbosity in the debug log, N – provides basic messages in the debug log.                                                                                                                                                                                                                                                                                                          |

#### 14.11.22.1 Running the Data Pool Cloud Cover Utility

- 1 Log in at the machine where the Data Pool Remove Collection Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Data Pool Remove Collection Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
- 3 At the prompt, type the command to start the Data Pool Cloud Utility, in the form of the following (perform only one command):

```
EcDICloudCoverUtilityStart <MODE> -operation
<correct|remove|populate|repopulate> -collection <ShortName.VersionId> [-
recovery no]
```

Data Pool Cloud Cover Utility usage examples:

- **EcDICloudCoverUtilityStart TS2 -operation remove -collection "MYD14.86"**

Remove all cloud cover entries for granules of collection "MYD14.86"

- **EcDICloudCoverUtilityStart TS2 -operation populate -collection "MYD29P1N.5"**

Populated all cloud cover sources for granules of collection "MUD29P1N.5"

- **EcDICloudCoverUtilityStart TS2 -operation repopulate -collection "MOD13A1.005"**

Update all cloud cover source for granules of collection "MOD13A1.005"

- **EcDICloudCoverUtilityStart TS2 -operation correct -collection "MOD29E1D.5"**

Check if the cloud cover source is correct given the DPM GUI configuration for collection "MOD29E1D.5", update the values to the correct ones if necessary.

### 14.11.23 Running the Data Pool Checksum Verification Service

The Data Pool Checksum Verification Service (CVS) provides a mechanism by which the ECS Operations Staff can perform checksum verification of files in the Data Pool. It runs as a background server process, and requests for file verification are submitted to it via the database using a utility script.

- The server process is highly configurable, and permits throttling of checksum operations by host and/or file system.
- The result of each verification will be logged to a special CVS log file (DICvChecksum.log) in the modes data/DPL directory. The information logged includes the time stamp, Data Pool granule ID, and the pass/fail state. In the case of a failure, the failure reason will also be logged, e.g file size mismatch, file missing, granule no longer exists, checksum value mismatch, etc.
- Upon successful checksum verification, it will update the last checksum verification timestamp in the AIM Inventory database for each file verified.
- Upon detection of a checksum verification failure, it will update the files checksum status to indicate the failure.

Table 14.11-26 provides a description of Command Line Parameters.

**Table 14.11-26. Command Line Parameter (1 of 2)**

| Parameter Name | Required | Description                                                                                   |
|----------------|----------|-----------------------------------------------------------------------------------------------|
| mode           | Yes      | Required parameter to specify the mode CVS is running.                                        |
| esdts          | No       | Optional parameter to specify ESDTs needs to be verified.                                     |
| granules g     | No       | Optional parameter to specify a list of Data Pool granules ids on the command line.           |
| filesystem fs  | No       | Optional parameter to specify Data Pool File system in which the granules are to be verified. |

**Table 14.11-26. Command Line Parameter (2 of 2)**

| Parameter Name | Required | Description                                                                                                                               |
|----------------|----------|-------------------------------------------------------------------------------------------------------------------------------------------|
| profile        | No       | Parameter to specify desired action after the file is checksum verified or checksum verification failed, such as email notification, etc. |
| file f         | No       | Parameter to specify a list of DataPool granule ids or files to be verified.                                                              |
| priority       | No       | Parameter to specify the priority of the granules so that if certain granules need to be checksummed before the others.                   |
| batchlabel     | No       | Parameter to specify the batch label associated with the batch of granules to be verified.                                                |
| help h         | No       | Parameter to specify the options to read the help page.                                                                                   |

### DataPool Checksum Server Configuration File

The DataPool Checksum Verification Service uses a configuration file: EcDIChecksumServer.CFG, located in /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table 14.11-27 describes the configuration parameters.

**Table 14.11-27. Configuration Parameters**

| Parameter Name              | Description                                                      |
|-----------------------------|------------------------------------------------------------------|
| DPL_USER                    | The user name for the Sybase connection.                         |
| DPL_SERVER                  | The name of the host Sybase SQL server is on.                    |
| DPL_PASSWDSEED              | The database password seed.                                      |
| DPLChecksumServiceHost Port | The port number of CVS.                                          |
| DPL_DATABASE                | The name of the AIM Inventory database you intend to connect to. |
| DPLChecksumServiceHosts     | CVS service hosts.                                               |
| DPL_MinPoolSize             | The minimal size of the DPL database connection pool.            |
| DPL_MaxPoolSize             | The maximum size of the DPL database connection pool.            |
| AppLogSize                  | The application log size.                                        |
| AppLogLevel                 | The application log level.                                       |
| DebugLevel                  | The debug log level.                                             |

#### 14.11.23.1 Running the Data Pool Checksum Verification Service Client

- 1 Log in at the machine where the Data Pool Checksum Verification Utility is installed (e.g., x4dpl01).
- 2 To change to the directory for starting the Data Pool Checksum Verification Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
  - The working directory is changed to **/usr/ecs/<MODE>/CUSTOM/utilities**.

- 3 At the prompt, type the command to start the Data Pool Checksum Verification Utility, in the form of the following.

```
EcDIInsertChecksumRequest.pl -mode <MODE> [-file|f filename] [-granule|g id] [-esdt esdtname] [-profile name] [-filesystem|fs dplfilesystem] [-priority priority] [-batchlabel batch] [-help|h]
```

Example of Data Pool Checksum Verification Service Client:

- **EcDIInsertChecksumRequest.pl -mode OPS -g 12345**  
Verify checksum for Data Pool granule 12345 in OPS mode
- **EcDIInsertChecksumRequest.pl -mode OPS -esdt AST\_L1A.004**  
Verify checksums for all granules from collection AST\_L1A.004 in OPS mode
- **EcDIInsertChecksumRequest.pl -mode OPS -file /home/cmshared/test.dat**  
Verify checksums for all granules specified in file /home/cmshared/test.dat

#### 14.11.24 Running the DPL XML Check Utility

The DPL XML Check utility provides a mechanism by which the ECS Operations Staff can periodically check for corruption of the XML files in datapool

In order to detect corruption, the utility verifies the contents of the files are well formed using xmllint.

The DPL XML Check utility should be started by the user cmshared (or similar). The XML Check utility is started by entering the following command:

```
EcDIXcu.pl <mode> <command line parameters>
```

There are seven command line parameters that may be used. Table 14.11-28 provides a description of those parameters.

**Table 14.11-28. Command Line Parameter**

| Parameter Name | Required                                             | Description                                                                                                                         |
|----------------|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| days           | No                                                   | Optional parameter to specify days since last checked.                                                                              |
| percent        | No                                                   | Optional parameter to specify percentage of files to check                                                                          |
| ESDT           | Yes, if granuleid or file parameters are not present | Parameter to specify which ESDTs to check. This is a comma separated list (no spaces). Can also specify "ALL" to include all ESDTs. |
| startdate      | No                                                   | Optional parameter used with –ESDT option. Specifies starting insert date to use for ESDTs.                                         |
| enddate        | No                                                   | Optional parameter used with –ESDT option. Specifies ending insert date to use for ESDTs.                                           |
| granuleid      | Yes, if ESDT or file parameters are not present      | Parameter to specify which granules to check. This is a comma separated list (no spaces).                                           |
| file           | Yes, if ESDT or granuleid parameters are not present | Parameter to specify which granules to check. Granule ids should be listing in a file separated by newlines.                        |
| outputDir      | No                                                   | Parameter to specify directory for error files under /workingdata/emd/<MODE>/DplXcu                                                 |

### DPL XML Check Configuration File

The DPL XML Check utility uses a configuration file, EcDI.CFG, located in /usr/ecs/<mode>/CUSTOM/cfg directory. The configuration parameters are stored in a PARAMETER = VALUE format with each parameter/value pair as a separate line entry in the file. Table 14.11.29 describes the configuration parameters.

**Table 14.11-29. Configuration Parameters**

| Parameter Name        | Description                                                                                                                                                          |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SYB_USER              | Sybase login name for the user of the Inventory database.                                                                                                            |
| SYB_SQL_SERVER        | Name of Sybase SQL Server hosting Inventory database.                                                                                                                |
| SYB_DBNAME            | Name of Inventory database.                                                                                                                                          |
| PGM_ID                | Program identifier used as seed to generate database password.                                                                                                       |
| NUM_RETRIES           | Number of times database operation will be attempted.                                                                                                                |
| RETRY_INTERVAL        | Number of seconds between retries.                                                                                                                                   |
| MAX_CONCUR_CHECKS     | Number of concurrent calls to xmllint that will be allowed.                                                                                                          |
| VALIDATION_OUTPUT_DIR | The default directory to place error output files. The directory should be readable/writeable by cmshared. The suggested directory is /workingdata/emd/<MODE>/DplXcu |

### 14.11.24.1 Running the DPL XML Check Utility

1 Log in at the machine where the DPL XML Check Utility is installed (e.g., x4oml01).

2 To change to the directory for starting the DPL XML Check Utility, type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.

3 At the prompt, type the command to start the DPL XML Check Utility, in the form of the following.

The DPL XML Check utility should be started by the user cmshared (or similar). The DPL XML Check utility is started by entering the following command:

**EcDlXcu.pl <mode> <command line parameters>**

XML Check Utility usage examples:

- For an "ESDT" run:

```
EcDlXcu.pl OPS -ESDT ALL
```

The DPL XML Check Utility will perform checking for all xml files in OPS mode

```
EcDlXcu.pl OPS -ESDT AST_L1A.003,MOD29.005 -startdate Jan 20 2008 -
enddate Dec 1 2008
```

The DPL XML Check Utility will performed checking for all AST\_L1A.003 and MOD29.005 xml files whose granules have been inserted between Jan 20 2008 and December 1 2008.

```
EcDlXcu.pl OPS -ESDT AST_L1B.003 -percent 50 -days 10
```

The DPL XML Check Utility will perform checking for 50% of AST\_L1B.003 granules which have not been checked in the last 10 days.

- For a "granuleid" run:

```
EcDlXcu.pl OPS -granuleid 22083,22085,22087
```

The DPL XML Check Utility will perform checking for the xml files related to granules 22083, 22085, and 22087 in OPS mode

- For a "file" run:

```
EcDlXcu.pl OPS -file granuleids.txt
```

The DPL XML Check Utility will perform checking for the xml files related to granules specified in granuleids.txt