

17. Archive Management/DataPool Maintenance

17.1 Archive Management Overview

Archive processing is at the heart of the system. Through archive processing, data that have been ingested into the system are archived to tape for permanent storage and distributed to users via hard media (tape or disk) or electronic means. The DAAC Archive Manager's job entails working with all levels of the system.

The Evolution Release 7.20 resulted in changes on the hardware and system levels for managing the the Archive. A migration from the AMASS-based archive to a StorNext SAN-based archive was implemented prior to and as a prerequisite of Release 7.20. A pair of HP570 G4 commodity-based Linux servers were installed to run the StorNext Storage Manager (SNSM) product in a high availability, failover environment.

The StorNext disk cache resides on SAN-attached-disk-storage managed by SNSM. This change benefits the DAACs with added capability for multi-tier storage management. StorNext Storage Manager (SNSM) is a hierarchical storage management (HSM) system for managing data on multiple storage tiers consisting of disk and tape resources. The purpose of SNSM in the ECS is to provide an easy-to-use interface (GUI based) for large data archives.

17.2 Archive Hardware

The archive hosts in the ECS architecture are a pair of Hewlett-Packard (HP) Proliant ML570 servers running Red Hat Linux 4.0 Update 3. SNSM includes a MetaData Controller (MDC) failover capability to support a higher level of availability than with the AMASS solution. 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 will significant increase the the uptime of the ECS archive at each DAAC. The archive uses one major type of archival storage hardware for storing science data, browse data, and other data. The StorageTek (STK) Powderhorn Model 9310 Automated Cartridge System tape storage tower, illustrated in Figure 2, is a mass storage system of removable media jukeboxes. The software that manages the storage in the system architecture is hosted on a HP Proliant ML570. The typical data storage archive consists of the following major elements: Library Storage Module (LSM), Powderhorn Model 9310, Automated Cartridge System (ACS) tape storage tower. Cartridge Access Port (CAP), where media are inserted or ejected from the LSM; standard capacity is 21 cartridges.

Dual tape-transport "robots" for moving cartridges from the tower to a tape drive or CAP and from the tape drive or CAP to the tower.

Tape drive rack with eight 9940 cartridge tape drives (rack capacity is 20 drives)

Library Management Unit (LMU), Model 9330, a serial port for the ACS Library Software (ACSL) that controls and monitors the ACS.

- Library Control Unit (LCU), Model 9311, a hardware interface for managing LSM intercommunications.

Each LSM tape archive can store thousands of tapes. The archive stores science data on STK 9940A and 9940B tapes, each 9940A tape can store 60 gigabytes of data (up to 200 gigabytes compressed) and each 9940B tape can store 200 gigabytes of data (up to 400 gigabytes compressed). Each 9940 tape cartridge is identified by a colored bar code label that shows the media number (see Figure 3). An archive catalog or database tracks the location of each cartridge within the library, based on information provided by the laser bar code reader

17.3 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, and to interpret them to the appropriate level needed to control and monitor archive operations. The StorNext Storage Manager (SNSM) software is a product of Quantum. Quantum, which acquired ADIC, is the vendor for both the StorNext File System and the StorNext Storage Management products.

The purpose of SNSM in the system is to provide an easy-to-use interface to a large tape archive. StorNext manages files, volumes (media), drives and jukeboxes. It allows UNIX File System (UFS) access methods to be employed (e.g., ftp, rcp, uucp, nfs, RPC, cp, mv and native commands) while removing some of the limitations of the UFS. Primary among these is reliance on 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.

17.4 Starting and Stopping StorNext

The ECS System contains both managed (Hericharchal 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 17.4-1 provides an Activity Checklist for Starting and Stopping StorNext.

Table 17.4-1. Starting and Stopping StorNext

Order	Role	Task	Section
1	System Administrator or Archive Manager	Starting the StorNext System	(P) 17.4.1.1 (P) 17.4.1.2
2	System Administrator or Archive Manager	Stopping the StorNext System	(P) 17.4.2.1
3	System Administrator or Archive Manager	Disabling the Tape Archive System	(P) 17.4.2.2
4	System Administrator or Archive Manager	Rebooting StorNext MedaData Server	(P) 17.4.3.1

17.4.1 Starting the StorNext Application

To bring the StorNext System, you must start both the server and its clients.

17.4.1.1 Start StorNext Manager Server (*this must be started first*) from the command line prompt: (as superuser or root)

-
- 1 Logon to the active metadata server (x4smvaa). Using x4smvaa will log you into the active (x4sml01 (primary) or x4sml02 (failover/secondary))
/etc/init.d/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)

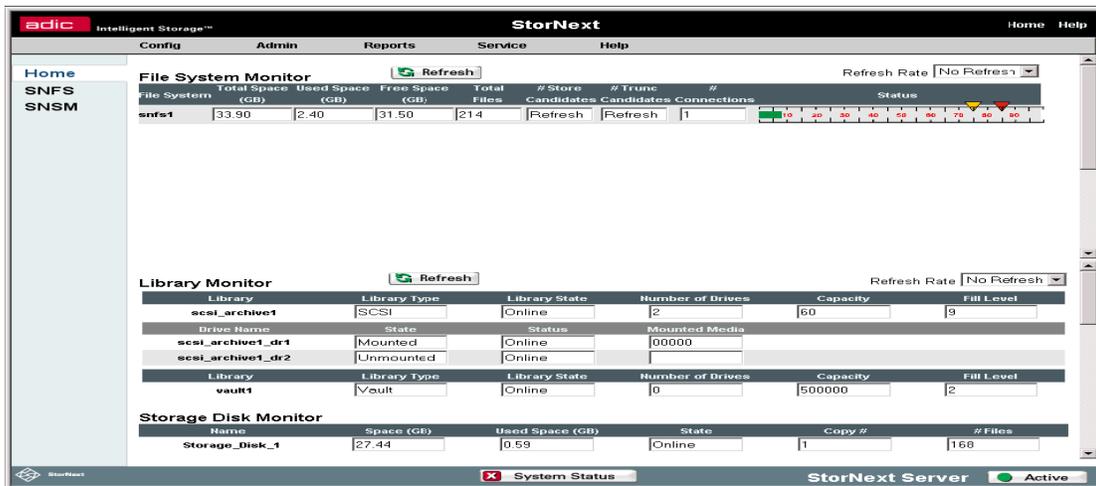
- 3 Logon to the each client hosts.
/etc/init.d/cvfs start
 - 4 Verify that the StorNext Manager Server is active with the command
#ps -ef | grep cvfs
-

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

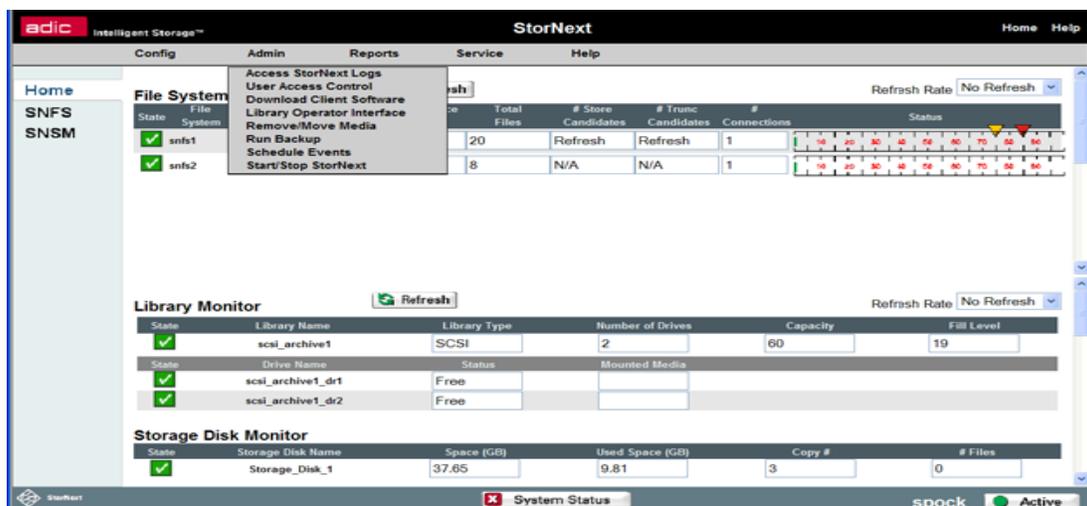
Note: Persons with Administrators Accounts can only have full control of the archive.

17.4.1.2 Start the StorNext Manager and its Clients from the GUI:

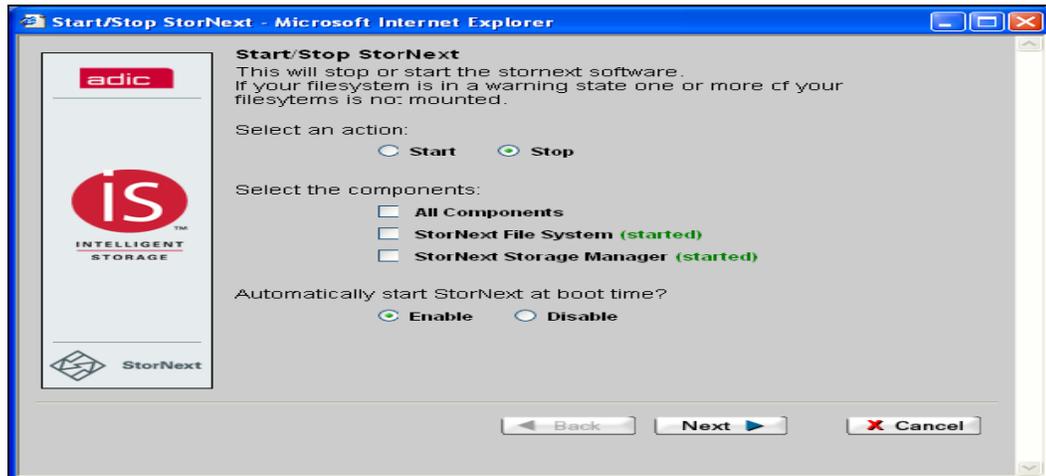
- 1 Bring up a Web browser. **Mozilla** is the EMD supported standard, however, other browsers such as Firefox, Netscape 7+ , and others can be used.
- 2 Enter the name of the active **StorNext** metadata server (p4smvaa:81).
 - The **StorNext** login window will appear.
- 3 Enter the username (admin) and password in the spaces provided. (Operators can create multiple accounts.)
 - The **StorNext** Home GUI will be displayed.



- 4 Select **Admin** from the Home Page.
 - The **Admin** Drop Down Menu will be displayed.



- 5 Select **Stop/Start StorNext** from **Admin** the pull-down window.
 - The **StorNext Start/Stop** page will be displayed.



- 6 Select the **Start** radio button for **Select an action**.
- 7 Select the **All Components** radio button for **Select the components** *”NOTE : NEVER Select Automatically Start StorNext Manager at boot time? EMD has provide a script in the init.d directory to perform this action.*
- 8 Select the **Next** radio button

17.4.2 Stopping the StorNext Application

To bring the StorNext System, you must stop both the server and and its clients. The Clients must all be stopped first.

17.4.2.1 Stop the StorNext Clients:

- 1 Log in as root (system administrator) into each **StorNext Metadata** client.
- 2 To stop the **StorNext** clients, type:
/etc/init.d/cvfs stop
- 3 Check to ensure client has been stopped
#ps -ef | grep cvfs

NOTE : MAKE SURE THAT ALL CLIENTS ARE STOPPED.

To Stop the **StorNext** server

4 Log in as **root** (system administrator) into the active **StorNext Metadata** server (x4smvaa).

5 To stop **StorNext** server, type:

```
# /etc/init.d/cvfs stop
```

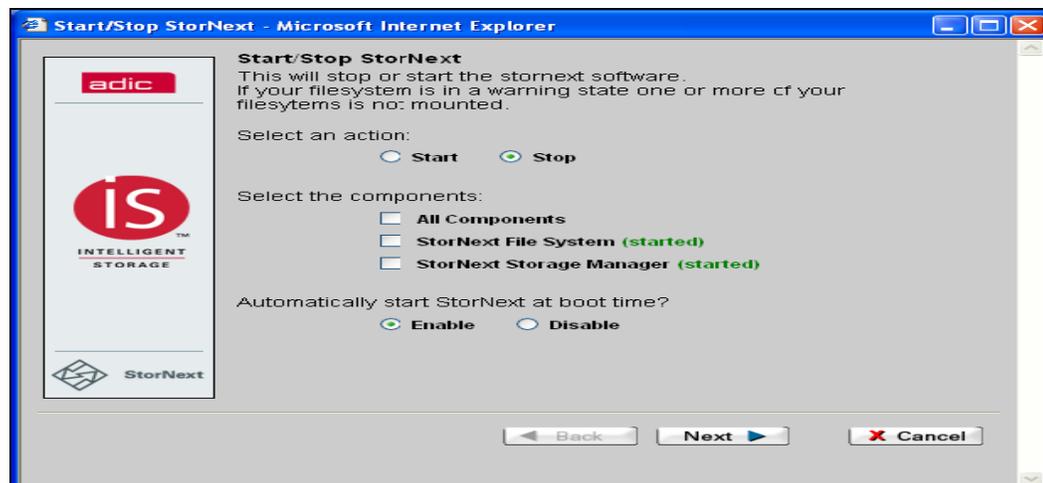
6 Check to ensure server has been stopped

```
#ps -ef | grep cvfs
```

17.4.2.2 Disabling the Tape Archive System

1 From the **StorNext Home Page**, select **Stop/Start StorNext** from **Admin Pull-down**.

- The **Stop/Start StorNext** page will be displayed.



2 Select the **Stop** radio button form the **Select an Action** section.

3 Select **StorNext Storage Manger** radio button form the **Select the components** section

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

4 Select the **Next** radio button

17.4.3 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 you must have root privileges. The following procedure demonstrates the steps to reboot StorNext MetaData Servers.

17.4.3.1 Rebooting the StorNext MetaData Servers

- 1 To reboot the **StorNext System**, you must stop both the Server and its Clients refer to Section 17.4.2 – **Stopping the StorNext Application**.
- 2 Perform Required Maintenance on StorNext Metadata Server.
- 3 Re-Start the StorNext Server and Clients refer to Section 17.4.1 – Starting the StorNext Application.

17.5 Loading and Removing Archive Media

For the STK storage facility, each Powderhorn is equipped with a 21-tape Cartridge Access Port (CAP). Tapes may be placed in the CAP for automatic loading. Tapes are also ejected through the CAP when identified for ejection using a command at the host for the STK Automated Cartridge System Library Software (ACSLs).

Table 17.5-1 provides an Activity Checklist for Loading, and Removing Archive Media.

Table 17.5-1. Loading and Removing Archive Media - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Manager	Loading Archive Media	(P) 17.5.1.1	
2	Archive Manager	Removing Archive Media	(P) 17.5.1.2.	

17.5.1 Loading Archive Media

Loading of media is appropriate when there are relatively small numbers of media to be loaded. Up to 21 volumes at a time may be loaded through the Cartridge Access Port (CAP). With automated loading, StorNext assigns each cartridge a unique volume number, enters the volumes in its database, and marks the volumes Online in the database.

17.5.1.1 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: **g**=GSFC, **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

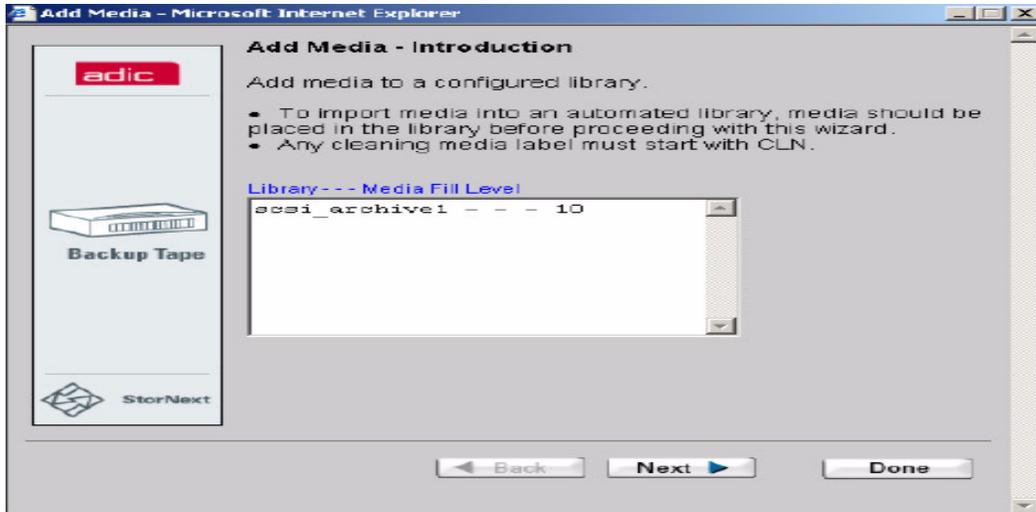
#vi /usr/adic/MSM/internal/config/media_file_”library”

Format :

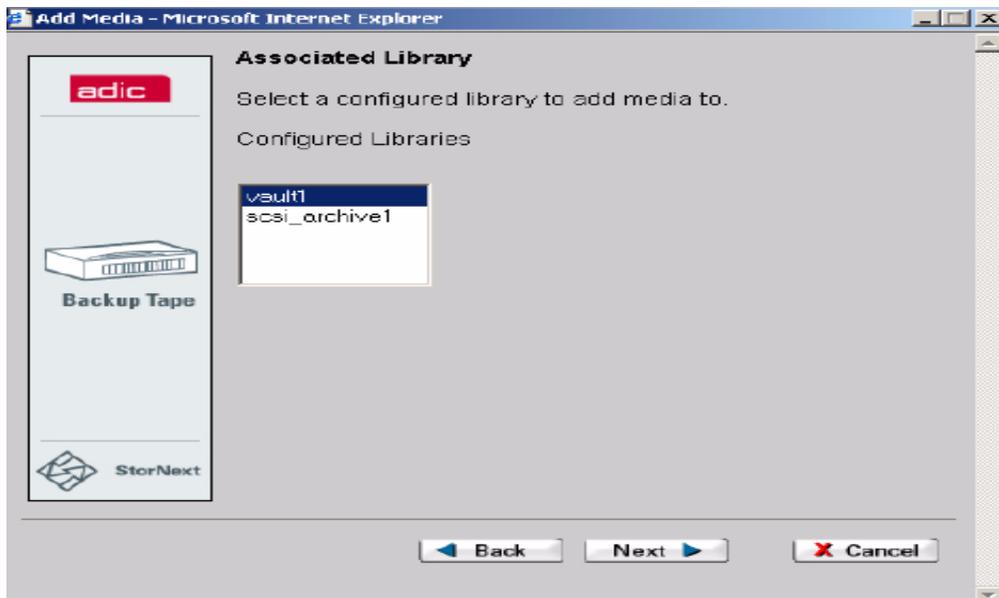
```
# [s]          any character in the set s, where s is a sequence of
#              characters and/or a range of characters, for example,
# [c-c].
#
# r*           zero or more successive occurrences of the regular
expression
#              r. The longest leftmost match is chosen.
# Examples:
#
# 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]

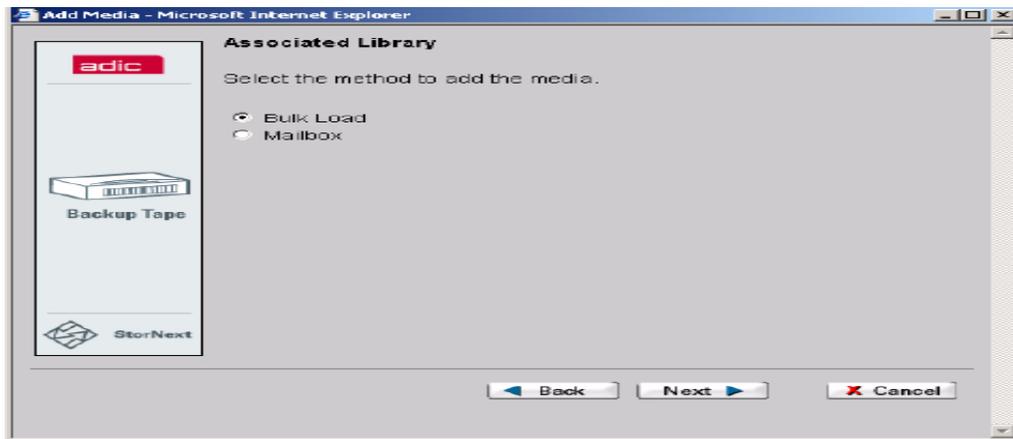
- 3 Place the Media in the Library. Select **Config-Add Media** from the **StorNext** Home page. The **Add Media – introduction** screen will appear. Select the appropriate library media, then select the **Next** button.



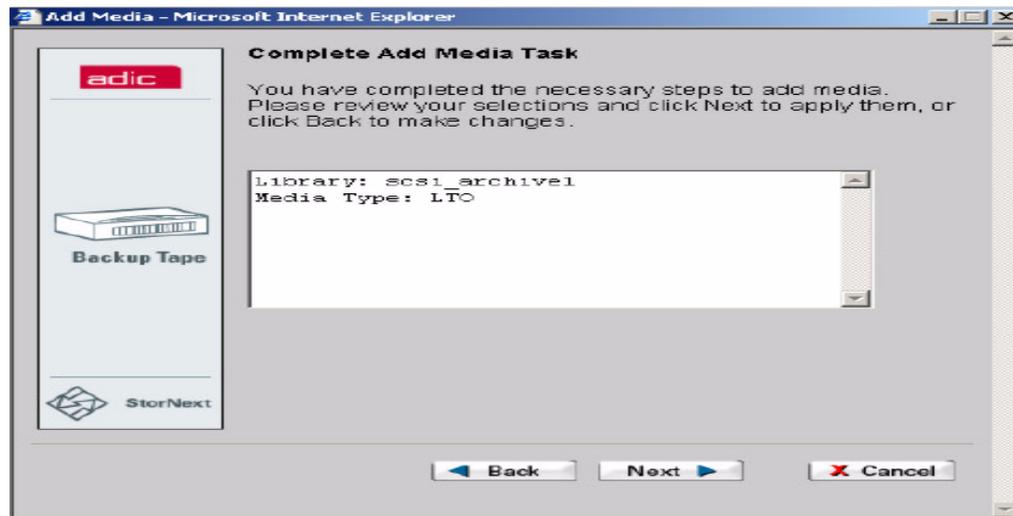
- 4 Select the **Associated Library**, and press the **Next** button.



- 5 Select the **Bulk Load** button from the **Associated Library** page, and press the **Next** button.



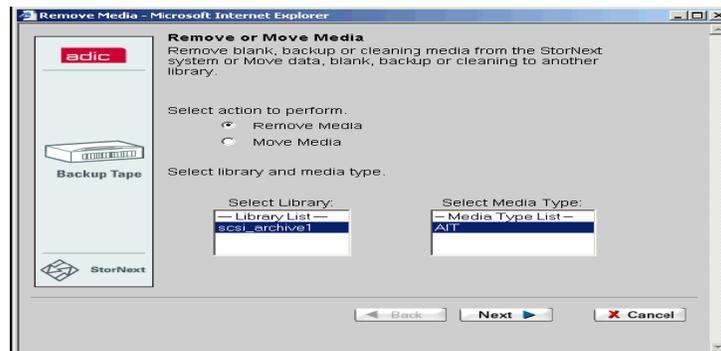
- 6 Select **Next** from the **Complete Add Media Task** screen.
 - The **Complete Add Media Task** screen displays.
 - The system will then automatically add your media



17.5.1.2 Removing Media

- 1 Select **Config-Remove** or **Move Media** from the **StorNext Home Page**. The **Remove Media or Move Media** screen will appear.
2. Select the **ReMove Media** button.

- 3 Select the appropriate **Library** and **Media Type**, then select the **Next** button.



- 4 Select the **Media** to be removed, then select the **Next** button.



- The **Completed Remove/Media Task** will appear.



- 5 When the status screen indicates that the media has been removed, select **Finish**. The **Library Operator Interface (LOL)** page will appear.
- 6 Open the recessed latch on the **Cartridge Access Port (CAP)** door and remove the tape(s).

7 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 exampe of what an entry in this file may look
like:
#AML_1 ESY2..
#
```

```
S2_98 SE925 [0-2]
```

17.6 Deleting Granules

The granule deletion capability provides the Operations Staff with the ability to delete granules using command line interfaces. The granules can be deleted from both the inventory and archive or just the archive.

The deletion process can involve deleting the specified SC (science) granules along with associated granules BR (browse), PH (production history), and QA (quality assurance) granules, as long as any other granules do not reference the associated granules. The deletion process can also involve deleting the specified granules even if they are inputs to other granules.

There are three phases to the granule deletion process:

- Logical deletion [marking granules (in the SDSRV database) for deletion].

- Physical deletion (removing from the SDSRV database data concerning granules marked for deletion and putting in the STMGT database data concerning granules to be deleted from the archive).
- Deletion from the archive (removing from the archive the files identified for deletion in the STMGT database).

Phase 1, Logical Deletion: 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 (SDSRV 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 SDSRV 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: The second phase is actual deletion from the inventory (SDSRV database) of the granules marked for physical deletion (not DFA only). Physical deletion occurs when the operations staff runs the Science Data Server Deletion Cleanup utility (EcDsDeletionCleanup.pl). The Deletion Cleanup utility removes all inventory rows (in the SDSRV database) for granules that were flagged as "deleted," including rows referencing related information (e.g., BR, PH, and QA), and the script creates entries in the DsStPendingDelete table (in the STMGT database) for granules to be deleted. This includes entries for granules that are to be physically deleted, as well as those designated DFA only.

Phase 3, Deletion from the Archive: In Phase 3 the operator uses the Archive Batch Delete utility to initiate the removal from the archive of the files listed in the DsStPendingDelete table (populated by SDSRV in Phase 2). The operator must run the utility separately on each appropriate Archive host (FSMS Server or APC server host) because the utility lists and deletes files from archive on the local host only. The Archive Batch Delete utility creates requests to the local archive server host to delete selected files. Files that are successfully deleted have their associated rows removed from the DsStPendingDelete table.

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 17.6-1 provides an Activity Checklist for Deleting Granules from the Archive.

Table 17.6-1. Deleting Granules - Activity Checklist

Order	Role	Task	Section
1	Archive Manager/Database Administrator	Generating a GeoID File to Be Used as Input to the Bulk Delete Utility or the Bulk Undelete Utility	(P) 17.6.1.1
2	Archive Manager/Database Administrator	Deleting Granules, Phase 1: Running the Bulk Delete Utility	(P) 17.6.2.1
3	Archive Manager/Database Administrator	Deleting Granules, Phase 2: Running the Deletion Cleanup Utility	(P) 17.6.3.1
4	Archive Manager/Database Administrator	Deleting Granules, Phase 3: Running the Archive Batch Delete Utility	(P) 17.6.4.1
5	Archive Manager/Database Administrator	“Undeleting” Granules from the Archive and Inventory	(P) 17.6.5.1

17.6.1 Generating a GeoID File to Be Used as Input to the Bulk Delete Utility or the Bulk Undelete Utility

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, 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.

NOTE: Finding 10,000 granules may take 15 to 20 minutes.

17.6.1.1 Generating a GeoID File to Be Used as Input to the Bulk Delete Utility or the Bulk Undelete Utility

-
- 1 Log in at the x4spl01 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 variables (if desired) at the UNIX prompt enter:

```
setenv MODE <ECS mode>
```

```
setenv SYB_SQL_SERVER <Sybase server>
```

```
setenv SYB_DB_NAME <database name>
```

- For example, the following commands would allow running the Bulk Search utility using the OPS mode Science Data server database at the ASDC (LaRC):

```
setenv MODE OPS
```

```
setenv SYB_SQL_SERVER I4dbI01_srvr
```

```
setenv SYB_DB_NAME EcDsScienceDataServer1
```

- 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 Science Data server 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., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_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 At the UNIX prompt enter:

```
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 <local granule ID file> ] -geoidfile <geoid file> [ -limit <granule limit> ] [ -mode <ECS mode> ] [ -server <Sybase server> ] [ -database <database name> ] -user <database login ID> [ -password <database login password> ]
```

- For example, the following command would generate a file of GeoIDs for all PM1ATTNR.077 granules marked “DFA” in the OPS mode at the ASDC (LaRC):

```
EcDsBulkSearch.pl -DFA -name PM1ATTNR -version 077 -mode OPS -server x4oml01_srvr -database EcDsScienceDataServer1 -user sdsrv_role -password password -geoidfile PM1ATTNR_Dec23.geoid
```

- If the **MODE**, **SYB_SQL_SERVER**, and **SYB_DB_NAME** environment variables had been set as shown in the example in Step 3, the following command would have the same effect as the previous example:

```
EcDsBulkSearch.pl -DFA -name PM1ATTNR -version 077 -user sdsrv_role -password password -geoidfile PM1ATTNR_Dec23.geoid
```

- The following command would generate a file of GeoIDs for all deleted (“physical” deletion) PM1ATTNR.077 granules in the OPS mode at the ASDC (LaRC):

```
EcDsBulkSearch.pl -physical -name PM1ATTNR -version 077 -user sdsrv_role -password password -geoidfile PM1ATTNR_Dec23.geoid
```

- The following command would generate a file of GeoIDs for all granules inserted in the OPS mode at the ASDC (LaRC) between 11:00 P.M. on May 6, 2005 and 6:00 P.M. on May 7, 2005:

```
EcDsBulkSearch.pl -insertbegin 05/06/2005 23:00:00 -insertend 05/07/2005 18:00:00 -mode OPS -server x4oml01_svr -database EcDsScienceDataServer1 -user sdsrv_role -password password -geoidfile 05062005GranList.geoid
```

- The **-name** option indicates that the search is restricted to granules with the specified ESDT ShortName only. Typically the **-name** option is used in conjunction with the **-version** option.
- The **-version** option indicates that the search is restricted to granules with the specified ESDT Version only. The **-version** option is used in conjunction with the **-name** option.
- The **-begindate** option indicates that the search is restricted to granules with a BeginningDateTime greater than or equal to the specified date (and optionally time). Typically the **-begindate** option is used in conjunction with the **-enddate** option.
- The **-enddate** option indicates that the search is restricted to granules with an EndingDateTime less than or equal to the specified date (and optionally time). Typically the **-enddate** option is used in conjunction with the **-begindate** option.
- The **-insertbegin** option indicates that the search is restricted to granules with an insertTime greater than or equal to the specified date (and optionally time). The **-insertbegin** option may be used in conjunction with the **-insertend** option.
- The **-insertend** option indicates that the search is restricted to granules with an insertTime less than or equal to the specified date (and optionally time). The **-insertend** option may be used in conjunction with the **-insertbegin** option.
- The **-acquirebegin** option indicates that the search is restricted to granules with a BeginningDateTime greater than or equal to the specified date (and optionally time). The **-acquirebegin** option is the same as the **-begindate** option except that it can be combined with the **-acquireend** option and used in a **BETWEEN** clause.
- The **-acquireend** option indicates that the search is restricted to granules with a BeginningDateTime less than or equal to the specified date (and optionally time).

Typically the **-acquireend** option is used in conjunction with the **-acquirebegin** option.

- The **-physical** option indicates that the search is restricted to deleted granules.
- The **-DFA** option indicates that the search is restricted to granules that are in a DFA (deleted from archive) status.
- The **<local granule ID file>** value specified for the **-localgranulefile** optional parameter identifies an input file containing Local Granule IDs that are to be converted into GeoIDs.
- The **<geoid file>** value specified for the **-geoidfile** parameter identifies the output file that contains the GeoIDs (e.g., **SC:PM1ATTNR.077:2013463060**) of the granules that meet the search criteria. A report file with the same name as the GeoID file but with a **.rpt** extension is generated for the purpose of verification.
- The **-limit** option indicates that the search should be limited to (return no more than) the specified number of granules.
- The **<ECS mode>** value specified for the **-mode** parameter indicates the ECS mode (e.g., OPS, TS1, or TS2) to be searched.
- The **<Sybase server>** value specified for the **-server** parameter indicates the Sybase (database) server (x4oml01_srvr) for the Science Data Server database.
- The **<database name>** value specified for the **-database** parameter indicates which database (e.g., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_TS2) is involved in the search.
- The **<database login ID>** value specified for the **-user** parameter is the user ID (e.g., sdsrv_role) for logging in to the Science Data Server database.
- The **<database login password>** value specified for the **-password** parameter is the password for logging in to the Science Data Server database. If the password option is not entered on the command line, the script prompts for a password.
- The Bulk Search utility runs and records GeoIDs of granules that meet the specified criteria in the GeoID file.

5 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 (e.g., SYN5_GranDel.geoid, Dec_23_AllDatatypes.geoid).
- For example:

```
x4oml01{cmops}[10]->vi PM1ATTNR_Dec23.geoid
SC:PM1ATTNR.077:2013463060
SC:PM1ATTNR.077:2013463061
SC:PM1ATTNR.077:2013463062
SC:PM1ATTNR.077:2013463063
SC:PM1ATTNR.077:2013463064
SC:PM1ATTNR.077:2013463065
SC:PM1ATTNR.077:2013463066
```

SC:PM1ATTNR.077:2013463067
SC:PM1ATTNR.077:2013463068
SC:PM1ATTNR.077:2013463069
[...]
SC:PM1ATTNR.077:2013496391
SC:PM1ATTNR.077:2013496392
SC:PM1ATTNR.077:2013496393
"PM1ATTNR_Dec23.geoid" 9740 lines, 255893 characters

– Many lines have been deleted from the example.

- Although this procedure has been written for the **vi** editor, any UNIX editor can be used to edit the file.

6 Review the file entries to identify 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.

7 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).

8 Press the **Esc** key.

9 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.
-

17.6.2 Deleting Granules, Phase 1: Running the Bulk Delete Utility

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:

- Physical.
- DFA.

A “physical” deletion run results in marking each specified granule and metadata as deleted from both inventory and archive. A “DFA” run involves marking each specified granule and metadata as deleted from the archive only.

As previously mentioned the Bulk Delete utility responds to operator-specified criteria for the deletion of granules by “logically” deleting from the inventory (SDSRV 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 SDSRV 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**.

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.)

NOTE: Deleting 10,000 granules may take 15 to 20 minutes.

17.6.2.1 Deleting Granules, Phase 1: Running the Bulk Delete Utility

1 Log in at the x4spl01 server host, where the Bulk Delete utility is installed.

2 To change to the directory for starting the Bulk Delete 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 <Sybase server>  
setenv SYB_DB_NAME <database name>
```

- For example, the following commands would allow running the Bulk Delete utility using the OPS mode Science Data Server database at the ASDC (LaRC):

```
setenv MODE OPS  
setenv SYB_SQL_SERVER x4oml01_srvr  
setenv SYB_DB_NAME EcDsScienceDataServer1
```

- The **<ECS mode>** value specified for the **MODE** parameter indicates the ECS mode (e.g., OPS, TS1, or TS2) in which to run the Bulk Delete utility.
 - If this environment variable is set, the **-mode** command line argument does not need to be given when starting the Bulk Delete utility.
- The **<Sybase server>** value specified for the **SYB_SQL_SERVER** parameter indicates the Sybase (database) server (e.g., x4oml01_srvr) for the Science Data Server database.
 - If this environment variable is set, the **-server** command line argument does not need to be given when starting the Bulk Delete utility.
- The **<database name>** value specified for the **SYB_DB_NAME** parameter indicates which database (e.g., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_TS2) is to be accessed by the Bulk Delete utility.
 - If this environment variable is set, the **-database** command line argument does not need to be given when starting the Bulk Delete utility.

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 perform “physical” deletion (i.e., deletion from both inventory and archive), at the UNIX prompt enter:

```
EcDsBulkDelete.pl -physical [ -delref ] [ -noassoc ] -geoidfile <geoid file> [ -mode <ECS mode> ] [ -server <Sybase server> ] [ -database <database name> ] -user <database login ID> [ -password <database login password> ] [ -log <log filename>]
```

- For example:

```
EcDsBulkDelete.pl -physical -geoidfile PM1ATTNR_Dec23.geoid -mode OPS -database EcDsScienceDataServer1 -user sdsrv_role -server x4oml01_srvr -password password -log PM1ATTNR_Dec23DelPhys.log
```

- The **-physical** option (i.e., delete from both inventory and archive) indicates that both granule metadata and archive are to be marked for deletion.
- 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.

NOTE: The **-delref** option has no effect on deleting SC/LM granules. They are always deleted regardless of whether or not they are referenced.

- The **-noassoc** option indicates that associated granules (e.g., QA, Browse, or PH granules) are not to be deleted.
- **<geoid file>** is the file that contains the GeoIDs of the granules to be deleted (in the form **SC:AST_L1BT.001:4267**).
- **<ECS mode>** is the ECS mode (e.g., OPS, TS1, or TS2) containing the granules to be deleted.
- **<Sybase server>** is the name of the applicable Sybase (database) server (e.g., x4oml01_srvr) for the Science Data Server database.
- **<database name>** is the name of the Science Data Server database (e.g., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_TS2) for the applicable mode.
- **<database login ID>** is the user ID (e.g., sdsrv_role) for logging in to the Science Data Server database.
- **<database login password>** is the database login ID’s password for logging in to the Science Data Server database. If the password option is not entered on the command line, the script prompts for a password.
- **<log filename>** is the name of the log file to which a deletion report is to be written. If no log file name is provided, the Bulk Delete utility creates a log file (BulkDelete.<date>.log) in the /usr/ecs/<MODE>/CUSTOM/logs/ directory.
- The Bulk Delete utility runs and deletes the granules specified in the GeoID file from the inventory and the archive.
- The Bulk Delete utility records information about utility events in the log file (e.g., PM1ATTNR_Dec23DelPhys.log).

5 To perform “DFA” deletion only, at the UNIX prompt enter:

```
EcDsBulkDelete.pl -DFA -geoidfile <geoid file> [ -mode <ECS mode> ] [ -server <Sybase server> ] [ -database <database name> ] -user <database login ID> [ -password <database login password> ] [ -log <log filename>]
```

- For example:

```
EcDsBulkDelete.pl -DFA -geoidfile PM1ATTNR_Dec23.geoid -mode OPS -database EcDsScienceDataServer1 -user sdsrv_role -server x4oml01_srvr -password password -log PM1ATTNR_Dec23DelDFA.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 runs and deletes the granules specified in the GeoID file from the archive.
- The Bulk Delete utility records information about utility events in the log file (e.g., PM1ATTNR_Dec23DelDFA.log).

6 When the Bulk Delete utility has completed its run and the log file is available, at the UNIX prompt enter:

```
pg <log filename>
```

- **<log filename>** refers to the log file to be reviewed (e.g., PM1ATTNR_Dec23DelPhys.log, PM1ATTNR_Dec23DelDFA.log).
- The first page of the log file is displayed.
- For example:

```
x4oml01{cmops}[10]->pg PM1ATTNR_Dec23DelPhys.log  
*****  
*           Bulk Granule Deletion Tool           *  
*                                                                 *  
*                                                                 *  
* Report created 23 Dec 2005 10:46:22                *  
*                                                                 *  
* GEOID file: /usr/ecs/OPS/CUSTOM/utilities/PM1ATTNR_Dec23.geoid *  
* Deletion mode: OPS                                *  
* Deletion type: physical                            *  
* Deleting associated granules: yes                  *  
*                                                                 *  
* Total granules processed: 9740                    *  
*****  
*                                                                 *  
* TOTAL UPDATED GRANULES                            *  
*                                                                 *  
* SC Type = 9705                                    *  
*                                                                 *  
* Associated granules deleted = 0                    *  
*                                                                 *
```

```

*****
*
* TOTAL FAILED GRANULES
*
* SC Type = 35
*
*****
SC:PM1ATTNR.077:2013463060 -- Granule successfully marked
SC:PM1ATTNR.077:2013463061 -- Granule successfully marked
SC:PM1ATTNR.077:2013463062 -- Granule successfully marked
SC:PM1ATTNR.077:2013463063 -- Granule successfully marked
SC:PM1ATTNR.077:2013463064 -- Granule successfully marked
SC:PM1ATTNR.077:2013463065 -- No matching granule found
SC:PM1ATTNR.077:2013463066 -- No matching granule found
SC:PM1ATTNR.077:2013463067 -- No matching granule found
SC:PM1ATTNR.077:2013463068 -- No matching granule found
SC:PM1ATTNR.077:2013463069 -- No matching granule found
[...]
SC:PM1ATTNR.077:2013496391 -- Granule successfully marked
SC:PM1ATTNR.077:2013496392 -- Granule successfully marked
SC:PM1ATTNR.077:2013496393 -- Granule successfully marked
(EOF):

```

– Many lines have been deleted from the example.

- Although this procedure has been written for the **pg** command, any UNIX editor or visualizing command (e.g., **vi**, **view**, **more**, **page**) can be used to review the file.

7 Review the log file to determine whether all deletions occurred as expected or experienced errors.

- The following **pg** commands (at the **:** prompt) are useful:
 - n then Return/Enter (go to Page n).
 - Return/Enter or +1 then Return/Enter (go down to the next page).
 - -1 then Return/Enter (go back to the preceding page).
 - +n then Return/Enter (go down n number of pages).
 - -n then Return/Enter (go back n number of pages).
 - +nl then Return/Enter (go down n number of lines).
 - -nl then Return/Enter (go back n number of lines).
 - \$ then Return/Enter [go to the last page (end of file)].
 - q then Return/Enter (exit from pg).
- The following types of messages related to granule states may be seen in the log file (as shown in the example in Step 6):
 - Granule successfully marked.
 - No matching granule found.
 - Cannot DFA this granule type.
 - Granule was logically deleted already.

- Golden granule cannot be deleted.
- Unknown error deleting granule.
- Granule NOT Deleted/DFAed Because Either It Is an Input To Another Granule OR It Is Associated With A Granule Which Is an Input To Another Granule.
- Granule was NOT locatable and therefore was NOT Deleted/DFAed.
- Granule was DFAed already.

8 If the GeoID and log files were written to the utilities directory (e.g., /usr/ecs/OPS/CUSTOM/utilities) rather than a home directory, log directory, or data directory, move or remove the files from the utilities directory.

- For example:

```
mv PM1ATTNR_Dec23DelPhys.log ../logs
mv PM1ATTNR_Dec23.geoid /home/cmops/geoid
```

- Leaving GeoID files and log files in the utilities directory can eventually make it more difficult to find needed scripts in the directory.
- The utilities directory is intended for scripts, not log files or data files. Leaving GeoID files and log files in the utilities directory causes it to load up with extraneous files.

17.6.3 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 SDSRV database) for granules that were flagged as “deleted,” including rows referencing related information (e.g., BR, PH, and QA), and the script creates entries in the DsStPendingDelete table (in the STMGT database) for granules to be deleted.

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.

17.6.3.1 Deleting Granules, Phase 2: Running the Deletion Cleanup Utility

- 1** Log in at the x4spl01 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 MY_MODE <ECS mode>
setenv SYB_USER sdsrv_role
```

```

setenv SYB_SQL_SERVER <Sybase server>
setenv SDSRV_DB_NAME <database name>
setenv STMGT_DB_NAME <STMGT database name>
setenv BATCH_SIZE <batch size>
setenv BATCH_SIZE_GRANULE <batch size granule>
setenv BATCH_SIZE_PH <batch size history>

```

- For example, the following commands would allow running the Deletion Cleanup utility using the OPS mode Science Data Server and Storage Management databases at the ASDC (LaRC):

```

setenv MY_MODE OPS
setenv SYB_USER sdsrv_role
setenv SYB_SQL_SERVER x4oml01_srvr
setenv SDSRV_DB_NAME EcDsScienceDataServer1
setenv STMGT_DB_NAME stmgtdb1
setenv BATCH_SIZE 5000
setenv BATCH_SIZE_GRANULE 30
setenv BATCH_SIZE_PH 5

```

- The <*ECS mode*> value specified for the **MY_MODE** parameter indicates the ECS mode (e.g., OPS, TS1, or TS2) in which to run the Deletion Cleanup utility.
- The <*Sybase server*> value specified for the **SYB_SQL_SERVER** parameter indicates the Sybase (database) server (e.g., x4oml01_srvr) for the Science Data Server database.
- The <*database name*> value specified for the **SDSRV_DB_NAME** parameter indicates which Science Data Server database (e.g., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_TS2) is to be accessed by the Deletion Cleanup utility.
- The <*STMGT database name*> value specified for the **STMGT_DB_NAME** parameter indicates which Storage Management database (e.g., stmgtdb1, stmgtdb1_TS1, or stmgtdb1_TS2) is to be accessed by the Deletion Cleanup utility.
- The <*batch size*> value specified for the **BATCH_SIZE** parameter indicates the size of the batch used to transfer granules from SDSRV to STMGT.
 - The default value is 10,000, which is accepted by pressing **Return/Enter** at the prompt without entering a value first.
- The <*batch size granule*> value specified for the **BATCH_SIZE_GRANULE** parameter indicates 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.
- The <*batch size history*> value specified for the **BATCH_SIZE_PH** parameter indicates the number of granules that will be deleted simultaneously from processing history tables during granule cleanup.
 - The default value is 5, which is accepted by pressing **Return/Enter** at the prompt without entering a value first.

- If the environment variables are set, the corresponding values do not have to be entered when starting the Deletion Cleanup utility.
- 4 To execute the Deletion Cleanup utility at the UNIX prompt enter:
EcDsDeletionCleanup.pl
- If the environment variable MY_MODE is **not** set, the utility prompts **Enter Mode of Operation :**.
- 5 If prompted, at the **Enter Mode of Operation :** prompt enter:
<MODE>
- If the environment variable SYB_USER is **not** set, the utility prompts **Enter Sybase User Name :**.
- 6 If prompted, at the **Enter Sybase User Name :** prompt enter:
sdsrv_role
- The utility prompts **Enter Sybase password:**.
- 7 At the **Enter Sybase password:** prompt enter:
<password>
- If the environment variable SYB_SQL_SERVER is **not** set, the utility prompts **Enter Sybase SQL Server Name :**.
- 8 If prompted, at the **Enter Sybase SQL Server Name :** prompt enter:
<Sybase server>
- **<Sybase server>** is the Sybase (database) server (e.g., x4oml01_srvr) for the Science Data Server database.
 - If the environment variable SDSRV_DB_NAME is **not** set, the utility prompts **Enter SDSRV's database name :**.
- 9 If prompted, at the **Enter SDSRV's database name :** prompt enter:
<SDSRV database name>
- The **<SDSRV database name>** is the relevant Science Data Server database (e.g., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_TS2).
 - If the environment variable STMGT_DB_NAME is **not** set, the utility prompts **Enter STMGT's database name :**.

- 10 If prompted, at the **Enter STMGT's database name :** prompt enter:
<STMGT database name>
- The *<STMGT database name>* is the relevant Storage Management database (e.g., stmgtdb1, stmgtdb1_TS1, or stmgtdb1_TS2).
 - If the environment variable BATCH_SIZE is **not** set, the utility prompts **Enter Batch Size for stmgt migration(10000) :**.

- 11 If prompted, at the **Enter Batch Size for stmgt migration(10000) :** prompt enter:
<batch size>
- *<batch size>* is the size of the batch used to transfer granules from SDSRV to STMGT.
 - The default value is 10,000, which is accepted by pressing **Return/Enter** at the prompt without entering a value first.
 - If the environment variable BATCH_SIZE_GRANULE is **not** set, the utility prompts **Enter Batch Size for Granule Deletion sdsrv(50) :**.

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.

- 12 If prompted, at the **Enter Batch Size for Granule Deletion sdsrv(50):** prompt enter:
<batch size granule>
- *<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.
 - If the environment variable BATCH_SIZE_PH is **not** set, the utility prompts **Enter Batch Size for Processing History Deletion sdsrv(5) :**.

- 13 If prompted, at the **Enter Batch Size for Processing History Deletion sdsrv(5):** prompt enter:
<batch size history>
- *<batch size history>* represents the number of granules that will be deleted simultaneously from processing history tables during granule cleanup.
 - The default value is 5, which is accepted by pressing **Return/Enter** at the prompt without entering a value first.
 - The utility prompts **Enter Log File name:**.

14 To use a log file name other than the default name at the **Enter Log File name:** prompt enter:

<log file name>

- *<log file name>* (e.g., **DeletionCleanup20051130.LOG**) is the name of the file where EcDsDeletionCleanup.pl records process events.
- To accept the default log file name (**DeletionCleanup.LOG**) press **Return/Enter** at the prompt without entering a value.
- If there is already a log file with the name that was entered, new entries will be appended to that log; otherwise a new log file will be created.
- If *<log file name>* does not include a **.LOG** file extension, Deletion Cleanup appends the extension to the name.
- Deletion Cleanup prepares to connect to the database and displays a **Ready to get into sdsrv database...** message.
- Then Deletion Cleanup connects to the database and checks for leftover granules that need to be processed. The following messages are displayed:

First check if there are leftover granules need to be processed*****

You may skip the following part if the log file for the previous run indicated a success or if it gave suggestion that it could be done automatically.

Do you want to proceed with this part? [y/n] :

15 To process the granules in the DsMdStagingTable only (to skip checking for leftover granules from a previous run of the Deletion Cleanup utility) at the **Do you want to proceed with this part? [y/n] :** prompt enter:

n

- The following message is displayed:
Operator chose to only process the granules left over in the DsMdStagingTable
- Deletion Cleanup either processes leftover granules or determines that there are no leftover granules from previous runs.
- Eventually the following message is displayed:
Do you wish to continue deleting some more granules? [y/n] :
- Go to Step 20.

16 To check for leftover granules from a previous run of the Deletion Cleanup utility at the **Do you want to proceed with this part? [y/n] :** prompt enter:

y

- The following message is displayed:
To ensure that all the granules are deleted for the previous run, StartTime, EndTime and subType file from the previous log file in the user input section need to be passed in.

If these info are not available in the log file, the previous run can not be recovered because script failed before the user input was gathered

Please enter the StartTime from the log file for the last run or hit return if not available:

- 17 To continue checking for leftover granules at the **Please enter the StartTime from the log file for the last run or hit return if not available:** prompt enter:

<start time>

- *<start time>* represents the start time (e.g., 10/15/2005) for the last run (from the log file).
- If the start time for the last run is not available, press **Return/Enter** at the prompt without entering a value.
- The utility prompts **Please enter the EndTime from the log file for the last run or hit return if not available:**

- 18 To continue checking for leftover granules at the **Please enter the EndTime from the log file for the last run or hit return if not available:** prompt enter:

<end time>

- *<end time>* represents the end time (e.g., 10/18/2005) for the last run (from the log file).
- If the end time for the last run is not available, press **Return/Enter** at the prompt without entering a value.
- The utility prompts **Please enter the subType file name from the log file for the last run or hit return if not available:**

- 19 To continue checking for leftover granules at the **Please enter the subType file name from the log file for the last run or hit return if not available:** prompt enter:

<subtype>

- *<subtype>* represents the subType (e.g., PM1ATTNR.077) for the last run (from the log file).
- If the subType for the last run is not available, press **Return/Enter** at the prompt without entering a value.
- Deletion Cleanup either processes leftover granules or determines that there are no leftover granules from previous runs.
- If there are leftover files, the following series of messages is displayed (there may be some time delay between messages):

Start processing leftover granules*****

executing...

No. of files to be migrated:: ...

Successfully Committed the migration of ...files

Successful Migration.

Finish processing leftover granules from previous run(s)*****

Do you wish to continue deleting some more granules? [y/n] :

- If there are no leftover files, the following series of messages is displayed:

No leftover granules from previous run(s)*****

Do you wish to continue deleting some more granules? [y/n] :

20 To exit Deletion Cleanup without doing any new cleanup at the **Do you wish to continue deleting some more granules? [y/n] :** prompt enter:

n

- An Operator chose to Terminate granule deletion process after processing the leftover granules message is displayed.
- Go to Step 30.

21 To proceed with deletion cleanup at the **Do you wish to continue deleting some more granules? [y/n] :** prompt enter:

y

- The following menu is displayed:

The user must select the menu option:

- 1. Select granules for a specific day (lag number or date <mm/dd/yyyy> format)**
 - 2. Select all granules older than (include) a specific day (lag number or date <mm/dd/yyyy> format)**
 - 3. QUIT**
- Choose 1, 2 or 3 ==>**

22 At the menu prompt enter:

<number>

- **<number>** is **1, 2,** or **3.**
- Enter **1** to clean up granules for a specific day.
 - The message **Enter Lag Time in Days or date <mm/dd/yyyy>:** is displayed.
- Enter **2** to clean up all granules older than (and including) a specific day.
 - The message **Enter Lag Time in Days or date <mm/dd/yyyy>:** is displayed.
- Enter **3** to exit from Deletion Cleanup.
 - The message **bye** is displayed.
 - Go to Step 30.

23 At the **Enter Lag Time in Days or date <mm/dd/yyyy>**: prompt enter:

<lag time>

- *<lag time>* is expressed as either a number of days (e.g., 17) or a date (in *mm/dd/yyyy* format, e.g., 10/18/2005).
 - An entry of **0** (zero) is equivalent to the current date.
- The lag time is the time from the current date at which the operator wishes to delete granules and is in units of days. For example, if the current date were November 30, 2005 and if it were necessary to delete all granules marked for deletion starting on November 25, 2005 (i.e. to delete all granules with a date of November 25, 2005 or earlier), the operator would enter either a lag time of **5** or the date **11/25/2005**.
- A lag time of **0** (zero) results in the deletion of **all** granules that have been flagged for deletion.
- One of the following types of messages is displayed:

**A Lag Time of '5' Days has been entered.
Is this correct? [y/n]**

OR

**A specific day of '11/25/2005' has been entered.
Is this correct? [y/n]**

24 If the displayed lag time is **not** correct, at the **Is this correct? [y/n]** prompt enter:

n

- An Enter Lag Time in Days or date <mm/dd/yyyy>: prompt message is displayed.
- Return to Step 23.

25 If the displayed lag time is correct, at the **Is this correct? [y/n]** prompt enter:

y

- The following menu is displayed:

The user must select the menu option:

1. Select above for specific datatypes(wildcard as an option) and versions

In the list of subtype format: AST_L1BT.001 or AST_L1B*.001...

2. Select the above for all datatypes and versions

3. QUIT

Choose 1, 2 or 3 ==>

26 At the menu prompt enter:

<number>

- *<number>* is **1**, **2**, or **3**.
- Enter **1** to specify a file of datatypes for cleanup.
 - The following message is displayed.

**This will select granules for specific datatypes in the User Specified Input File
Please specify the full path/filename ==>**

- Enter **2** to select all datatypes for cleanup.

– The following type of message is displayed:

Search against DsMdDeletedGranules table return the following results:

ShortName: PM1ATTNR, VersionID: 77, Num of Granules: 9740, DFA granules: YES

The number of Inventory Granules to be deleted = 9740

The number of DFA Granules to be deleted = 9740

Do you wish to continue deleting these granules?

- Enter **3** to exit from Deletion Cleanup.

– The message bye is displayed.

– Go to Step 30.

27 If a **Please specify the full path/filename ==>** prompt is displayed, enter:

<path>

- <path> is the full path and file name for the input file of data types for cleanup.
- For example: **/usr/ecs/OPS/CUSTOM/data/DSS/datatypes20051130.dat**
- The following type of message is displayed:

Search against DsMdDeletedGranules table return the following results:

ShortName: PM1ATTNR, VersionID: 77, Num of Granules: 9740, DFA granules: YES

The number of Inventory Granules to be deleted = 9740

The number of DFA Granules to be deleted = 9740

Do you wish to continue deleting these granules?

28 To **not** continue with the cleanup at the **Do you wish to continue deleting these granules?** prompt enter:

n

- The message Operator chose to Terminate granule deletion process after getting the counts from DsMdDeletedGranules is displayed.
- Go to Step 30.

29 To continue with the cleanup at the **Do you wish to continue deleting these granules?** prompt enter:

y

- The following series of messages is displayed (there may be some time delay between messages):

executing exec ProcPhysicalDeleteFmStaging 50,5

First delete all the science related granules

Begin Insert to #t1 for SC granules that will be deleted

....

.....

.....

finished Physical Deleting. Continue migrating deleted granules to Stmgt

No. of files to be migrated:: 2

Successful Migration.

Wed Nov 30 13:03:16 EST 2005

Successfully finished running Deletion CleanUp Utility.

30 To rerun the utility (e.g., in response to a message recommending that the utility be rerun) return to Step 4.

17.6.4 Deleting Granules, Phase 3: Running the Archive Batch Delete Utility

The operator uses the Archive Batch Delete utility to initiate the removal from the archive of the files listed in the DsStPendingDelete table (populated by SDSRV in Phase 2). The operator must run the utility separately on each appropriate Archive host (FSMS Server or APC Server host) because the utility lists and deletes files from archive on the local host only. The Archive Batch Delete utility creates requests to the local archive server host to delete selected files. Files that are successfully deleted have their associated rows removed from the DsStPendingDelete table.

17.6.4.1 Deleting Granules, Phase 3: Running the Archive Batch Delete Utility

- 1 Log in at the appropriate FSMS Server where the Archive Batch Delete utility is installed.
 - Examples of FSMS Server host names include e4smvaa, l4smvaa, and n4smvaa.
 - For detailed instructions refer to the procedure for Logging in to System Hosts (Section 17.3.1).
- 2 To change to the directory for starting the Archive Batch Delete 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 <Sybase server>
setenv SYB_DB_NAME <database name>
```

- For example, the following commands would allow running the Archive Batch Delete utility using the OPS mode Storage Management database at the ASDC (LaRC):

```
setenv MODE OPS
setenv SYB_SQL_SERVER x4oml01_srvr
setenv SYB_DB_NAME stmgtdb1
```

- The <ECS mode> value specified for the **MODE** parameter indicates the ECS mode (e.g., OPS, TS1, or TS2) in which to run the Archive Batch Delete utility.
- The <Sybase server> value specified for the **SYB_SQL_SERVER** parameter indicates the Sybase (database) server (e.g., x4oml01_srvr) for the Science Data Server database.
- The <database name> value specified for the **SYB_DB_NAME** parameter indicates which Storage Management database (e.g., stmgtdb1, stmgtdb1_TS1, or stmgtdb1_TS2) is to be accessed by the Archive Batch Delete utility.
- If the environment variables are set, the corresponding values do not have to be entered when starting the Archive Batch Delete utility.

4 To execute the Archive Batch Delete utility at the UNIX prompt enter:

```
EcDsStArchiveBatchDelete.pl
```

- If the environment variable **MODE** is **not** set, the utility prompts **Enter Mode of Operation :**.

NOTE: If **Ctrl-c** is pressed during the deletion process, the process is terminated, the utility prints out the reports and displays the message **Process was aborted**.

5 If prompted, at the **Enter Mode of Operation :** prompt enter:

```
<MODE>
```

- The utility prompts **Enter log file name (or return if use default log file):**.

6 To use a log file name other than the default name at the **Enter log file name (or return if use default log file):** prompt enter:

```
<log file name>
```

- <log file name> (e.g., **EcDsStArchiveBatchDelete.ALOG**) is the name of the file where **EcDsStArchiveBatchDelete.pl** records process events.
- To accept the default log file name (**ArchiveBatchDelete.<date>.<time>.log**) press **Return/Enter** at the prompt without entering a value.

- The utility displays a message similar to the following message:
log information will be put to
/usr/ecs/OPS/CUSTOM/logs/ArchiveBatchDelete.20051201.161030.log
- The utility prompts **Enter db login :**

7 At the **Enter db login :** prompt enter:
stmgt_role

- The utility prompts **Enter db password :**

8 At the **Enter db password :** prompt enter:
<password>

- If the environment variable SYB_SQL_SERVER is **not** set, the utility prompts **Enter Sybase SQL Server Name :**

9 If prompted, at the **Enter Sybase SQL Server Name :** prompt enter:
<Sybase server>

- *<Sybase server>* is the Sybase (database) server (e.g., x4oml01_srvr) for the Storage Management database.
- If the environment variable SYB_DB_NAME is **not** set, the utility prompts **Enter STMGT's database name :**

10 If prompted, at the **Enter STMGT's database name :** prompt enter:
<database name>

- The *<database name>* is the relevant Storage Management database (e.g., stmgtdb1, stmgtdb1_TS1, or stmgtdb1_TS2).
- Archive Batch Delete connects to the database, lists MissingVolGroup information from the Stage column of the DsStPendingDelete table, and allows the operator to reset the Stage field to NULL. It displays the following kinds of messages:

Ready to get into stmgt database... stmgtdb1.

Listing MissingVolGroup information in DsStPendingDelete...

The following is the list of MissingVolGroup information in DsStPendingDelete table:

VersionedDataType	filecount
--------------------------	------------------

total file count: 0

No files in DsStPendingDelete table having stage MissingVolGroup.

Do you want to continue? [y/n] :

- In the example no file was found with missing volgroup.

- 11 To exit from the utility at the **Do you want to continue? [y/n] :** prompt enter:
n
- The following message is displayed:
Terminating batch delete process.
- 12 To continue with the archive batch deletion process at the **Do you want to continue? [y/n] :** prompt enter:
y
- The following types of messages are displayed:
Trying to get HardWareCI from stmgtdb1...
HardWareCI is DRP1
server id is 3
exec DsStPDInsertComplete to fill on DsStPendingDelete, this may take a while...
 - Then there is a prompt for the batch size for log deletion:
Please enter batch size for log deletion progress (or return if use default value 100) :
- 13 To continue with the archive batch deletion process at the **Please enter batch size for log deletion progress (or return if use default value 100) :** prompt enter:
<batch size>
- <batch size>** represents the batch size for log deletion progress, which means that every time the specified number of files have been processed Archive Batch Deletion reports/logs the total number of files successfully deleted and total number of files which failed to be deleted.
 - To accept the default value (100) press **Return/Enter** at the prompt without entering a value.
 - If the default value (100) is accepted, every time 100 files have been processed Archive Batch Deletion reports/logs the total number of files successfully deleted and total number of files which failed to be deleted.
 - The utility shows deletion information from the DsStPendingDelete table:
Listing information in the DsStPendingDelete...
- The following is the list of information in DsStPendingDelete table for DRP1 :
- | | VersionedDataType | VolumeGroupId | VolumeGroupPath | filecount |
|----|-------------------|---------------|-------------------------|-----------|
| 0: | AST_L1BT.001 | 182 | /archive/OPS/drp/astl1 | 4 |
| 1: | AST_L1BT.001B | 167 | /archive/OPS/acm/browse | 4 |
| 2: | Browse.001 | 104 | /archive/OPS/acm/browse | 2 |
- total file count: 10**
- Do you want to continue? [y/n] :**

14 To exit from the utility at the **Do you want to continue? [y/n]** : prompt enter:

n

- The following message is displayed:

Terminating batch delete process.

15 To continue with the archive batch deletion process at the **Do you want to continue? [y/n]** : prompt enter:

y

- Then there is a prompt for selecting the index number(s) of the VersionedDataType/VolumeGroupPath set(s) to be deleted:

Please enter the indexes of the desired VersionedDataType and VolumeGroupPath pair with space between indexes or just -1 for all)=>

16 To continue with the archive batch deletion process at the **Please enter the indexes of the desired VersionedDataType and VolumeGroupPath pair with space between indexes or just -1 for all)=>** prompt enter:

<index1> [<index2> ... <indexn>]

- **<index1> [<index2> ... <indexn>]** represent the index numbers (from the previously displayed list) of the VersionedDataType/VolumeGroupPath sets to be deleted. (Refer to the example in Step 13 for a list of VersionedDataType/VolumeGroupPath sets with index numbers.)
- To select all VersionedDataType/VolumeGroupPath sets for deletion either enter all index numbers separated by spaces or just enter **-1**.
- For example:

Please enter the indexes of the desired VersionedDataType and VolumeGroupPath pair with space between indexes or just -1 for all)=> 0

You selected the following VersionedDataType/VolumeGroupPath sets:

0: AST_L1BT.001 182 /archive/OPS/drp/astl1 4

- In the example index number **0** was entered and Archive Batch Deletion displayed relevant information associated with that entry in the DsStPendingDelete table (as shown in Step 13)

- Then there is a prompt for deleting a subset of the selected granules:

Do you want to delete a subset of the granules above? [y/n] :

17 To delete the entire set of selected granules (to skip selecting a subset) at the **Do you want to delete a subset of the granules above?** [y/n] : prompt enter:

n

- The following types of messages are displayed:

Select files in DsStPendingDelete for deletion...

Start of Batch Delete...

get Directory /archive/OPS/drp/astl1

Total No. of files to be deleted in this session: 3

No. of files to be deleted: 0, deleted: 3, failed to be deleted: 1

Files deleted in this session: 3

Files failed to be deleted in this session: 1

Total files deleted so far: 3

Total files failed to be deleted so far: 1

Listing information in the DsStPendingDelete...

The following is the list of information in DsStPendingDelete table for DRP1 :

	VersionedDataType	VolumeGroupId	VolumeGroupPath	filecount
0:	AST_L1BT.001	182	/archive/OPS/drp/astl1	1
1:	AST_L1BT.001B	167	/archive/OPS/acm/browse	4
2:	Browse.001	104	/archive/OPS/acm/browse	2

total file count: 7

Do you want to continue? [y/n] :

- The example indicates that three files were deleted from Index 0 (AST_L1BT.001) and one failed to be deleted. (Compare with the example in Step 13.)
- Return to Step 14.

18 To delete a subset of the selected granules at the **Do you want to delete a subset of the granules above?** [y/n] : prompt enter:

y

- Then there is a prompt for entering the subset size:

Please enter subset size :

- If a subset is entered, the utility selects the file(s) to be deleted based on the selected VersionedDataType/VolumeGroupPath and the granule FileName. Older files are deleted first.

19 To delete a subset of the selected granules at the **Please enter subset size :** prompt enter:

<subset size>

- *<subset size>* represents the subset size.
- For example:

Please enter subset size : 1

Subset size: 1

Select files in DsStPendingDelete for deletion...

Start of Batch Delete...

get Directory /archive/OPS/drp/astl1

Total No. of files to be deleted in this session: 1

No. of files to be deleted: 0, deleted: 1, failed to be deleted: 0

Files deleted in this session: 1

Files failed to be deleted in this session: 0

Total files deleted so far: 1

Total files failed to be deleted so far: 0

Listing information in the DsStPendingDelete...

The following is the list of information in DsStPendingDelete table for DRP1 :

	VersionedDataType	VolumeGroupId	VolumeGroupPath	filecount
0:	AST_L1BT.001	182	/archive/OPS/drp/astl1	3
1:	AST_L1BT.001B	167	/archive/OPS/acm/browse	4
2:	Browse.001	104	/archive/OPS/acm/browse	2

total file count: 9

Do you want to continue? [y/n] :

- The example indicates that one file was deleted from Index 0 (AST_L1BT.001). (Compare with the example in Step 13.)

20 Return to Step14.

17.6.5 “Undeleting” Granules from the Archive and Inventory

In the event that it is desirable to restore granules that have been marked for deletion in the Science Data Server database (e.g., using the Bulk Delete utility), the Bulk Undelete utility provides an “undelete” capability.

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.)

NOTE: “Undeleting” 10,000 granules may take 15 to 20 minutes.

17.6.5.1 “Undeleting” Granules from the Archive and Inventory

- 1 Log in at the Sun Consolidation Internal server host, where the Bulk Undelete utility is installed.
 - Examples of Sun Consolidation Internal server host names include **x4oml01**.
 - For detailed instructions refer to the procedure for **Logging in to System Hosts**

- 2 To change to the directory for starting the Bulk Undelete 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 <Sybase server>
setenv SYB_DB_NAME <database name>
 - For example, the following commands would allow running the Bulk Undelete utility using the OPS mode Science Data Server database at the ASDC (LaRC):
setenv MODE OPS
setenv SYB_SQL_SERVER x4oml01_srvr
setenv SYB_DB_NAME EcDsScienceDataServer1
 - The **<ECS mode>** value specified for the **MODE** parameter indicates the ECS mode (e.g., OPS, TS1, or TS2) in which to run the Bulk Undelete utility.
 - If this environment variable is set, the **-mode** command line argument does not need to be given when starting the Bulk Undelete utility.
 - The **<Sybase server>** value specified for the **SYB_SQL_SERVER** parameter indicates the Sybase (database) server (e.g., x4oml01_srvr) for the Science Data Server database.
 - If this environment variable is set, the **-server** command line argument does not need to be given when starting the Bulk Undelete utility.
 - The **<database name>** value specified for the **SYB_DB_NAME** parameter indicates which database (e.g., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_TS2) is to be accessed by the Bulk Undelete utility.
 - If this environment variable is set, the **-database** command line argument does not need to be given when starting the Bulk Undelete utility.

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:

```
EcDsBulkUndelete.pl -physical [ -noassoc ] -geoidfile <geoid file> [ -mode <ECS mode> ] [ -server <Sybase server> ] [ -database <database name> ] -user <database login ID> [ -password <database login password> ] [ -log <log filename>]
```

- For example:

```
EcDsBulkUndelete.pl -physical -geoidfile PM1ATTNR_Dec23.geoid -mode OPS -database EcDsScienceDataServer1 -user sdsrv_role -server x4oml01_srvr -password password -log PM1ATTNR_Dec23UndelPhys.log
```

- The **-physical** option (i.e., “undelete” from both inventory and archive) indicates that both granule metadata and archive are to be marked for “undeletion.”
- The **-noassoc** option indicates that associated granules (e.g., QA, Browse, or PH granules) are not to be “undeleted.”
- **<geoid file>** The file that contains the GeoIDs of the granules to be “undeleted” (in the form **SC:AST_L1BT.001:4267**).
- **<ECS mode>** is the ECS mode (e.g., OPS, TS1, or TS2) containing the granules to be “undeleted.”
- **<Sybase server>** is the name of the applicable Sybase (database) server (e.g., x4oml01_srvr) for the Science Data Server database.
- **<database name>** is the name of the Science Data Server database (e.g., EcDsScienceDataServer1, EcDsScienceDataServer1_TS1, or EcDsScienceDataServer1_TS2) for the applicable mode.
- **<database login ID>** is the user ID (e.g., sdsrv_role) for logging in to the Science Data Server database.
- **<database login password>** is the database login ID’s password for logging in to the Science Data Server database. If the password option is not entered on the command line, the utility prompts for a password.
- **<log filename>** is the name of the log file to which a deletion report is to be written. If no log file name is provided, the Bulk Undelete utility creates a log file (BulkUndelete.<date>.log) in the /usr/ecs/<MODE>/CUSTOM/logs/ directory.
- The Bulk Undelete utility runs and removes “deleted” markings for the granules/metadata specified in the GeoID file in both inventory and archive.
- The Bulk Undelete utility records information about utility events in the log file (e.g., PM1ATTNR_Dec23UndelPhys.log).

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

```
EcDsBulkUndelete.pl -DFA -geoidfile <geoid file> [ -mode <ECS mode> ] [ -server <Sybase server> ] [ -database <database name> ] -user <database login ID> [ -password <database login password> ] [ -log <log filename>]
```

- For example:
EcDsBulkUndelete.pl -DFA -geoidfile PM1ATTNR_Dec23.geoid -mode OPS -database EcDsScienceDataServer1 -user sdsrv_role -server x4oml01_srvr -password password -log PM1ATTNR_Dec23UndelPhys.log
- The **-DFA** option indicates that “deleted” markings are to be removed for the granules in the archive only.
- 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 (e.g., PM1ATTNR_Dec23UndelPhys.log).

6 When the Bulk Undelete utility has completed its run and the log file is available, at the UNIX prompt enter:

pg <log filename>

- **<log filename>** refers to the log file to be reviewed (e.g., PM1ATTNR_Dec23DelPhys.log, PM1ATTNR_Dec23DelDFA.log).
- The first page of the log file is displayed.
- For example:

```
x4oml01{cmops}[10]->pg PM1ATTNR_Dec23DelPhys.log
*****
*           Bulk Granule Undeletion Tool
*
*
* Report created 23 Dec 2005 10:46:22
*
* GEOID file: /usr/ecs/OPS/CUSTOM/utilities/PM1ATTNR_Dec23.geoid
* Undeletion mode: OPS
* Undeletion type: physical
* Undeleting associated granules: yes
*
* Total granules processed: 9740
*****
* TOTAL UPDATED GRANULES
*
* SC Type = 9740
*
* Associated granules undeleted = 4163
*
*****
*
* TOTAL FAILED GRANULES
*
*
*****
```

SC:PM1ATTNR.077:2013463060 -- Granule successfully marked
SC:PM1ATTNR.077:2013463061 -- Granule successfully marked
SC:PM1ATTNR.077:2013463062 -- Granule successfully marked
SC:PM1ATTNR.077:2013463063 -- Granule successfully marked
SC:PM1ATTNR.077:2013463064 -- Granule successfully marked
SC:PM1ATTNR.077:2013463065 -- Granule successfully marked
SC:PM1ATTNR.077:2013463071 -- Granule successfully marked
SC:PM1ATTNR.077:2013463072 -- Granule successfully marked
SC:PM1ATTNR.077:2013463073 -- Granule successfully marked
SC:PM1ATTNR.077:2013463074 -- Granule successfully marked
 [...]

SC:PM1ATTNR.077:2013496391 -- Granule successfully marked
SC:PM1ATTNR.077:2013496392 -- Granule successfully marked
SC:PM1ATTNR.077:2013496393 -- Granule successfully marked
(EOF):

- Many lines have been deleted from the example.

- Although this procedure has been written for the **pg** command, any UNIX editor or visualizing command (e.g., **vi**, **view**, **more**, **page**) can be used to review the file.

7 Review the log file to determine whether all “undeletions” occurred as expected or experienced errors.

- The following **pg** commands (at the **:** prompt) are useful:
 - **n** then Return/Enter (go to Page **n**).
 - Return/Enter or **+1** then Return/Enter (go down to the next page).
 - **-1** then Return/Enter (go back to the preceding page).
 - **+n** then Return/Enter (go down **n** number of pages).
 - **-n** then Return/Enter (go back **n** number of pages).
 - **+nl** then Return/Enter (go down **n** number of lines).
 - **-nl** then Return/Enter (go back **n** number of lines).
 - **\$** then Return/Enter [go to the last page (end of file)].
 - **q** then Return/Enter (exit from **pg**).
- The following types of messages related to granule states may be seen in the log file (as shown in the example in Step 6):
 - Granule successfully marked.
 - No matching granule found.
 - Cannot UnDFA this granule type.
 - Granule is not DFA'd.
 - Granule is not Logically Deleted.
 - Unknown error undeleting granule.
 - Granule no longer in DsMdDeletedGranules table.
 - Invalid UR format.

8 If the GeoID and log files were written to the utilities directory (e.g., /usr/ecs/OPS/CUSTOM/utilities) rather than a home directory, log directory, or data directory, move or remove the files from the utilities directory.

- For example:

```
mv PM1ATTNR_Dec23UndelPhys.log ../logs
mv PM1ATTNR_Dec23.geoid /home/cmops/geoid
```

- Leaving GeoID files and log files in the utilities directory can eventually make it more difficult to find needed scripts in the directory.
 - The utilities directory is intended for scripts, not log files or data files. Leaving GeoID files and log files in the utilities directory causes it to load up with extraneous files.
-

17.7 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. EMD SUSTAINING ENGINEERING DOES NOT SUPPORT ANY SYTEM 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. This backup includes:

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

A partial backup is run on all other days of the week (that the full backup is not run). This backup includes:

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

Table 17.7-1 provides an Activity Checklist for StorNext Backup procedures addressed in this section.

Table 17.7-1. StorNext Backup Procedures - Activity Checklist

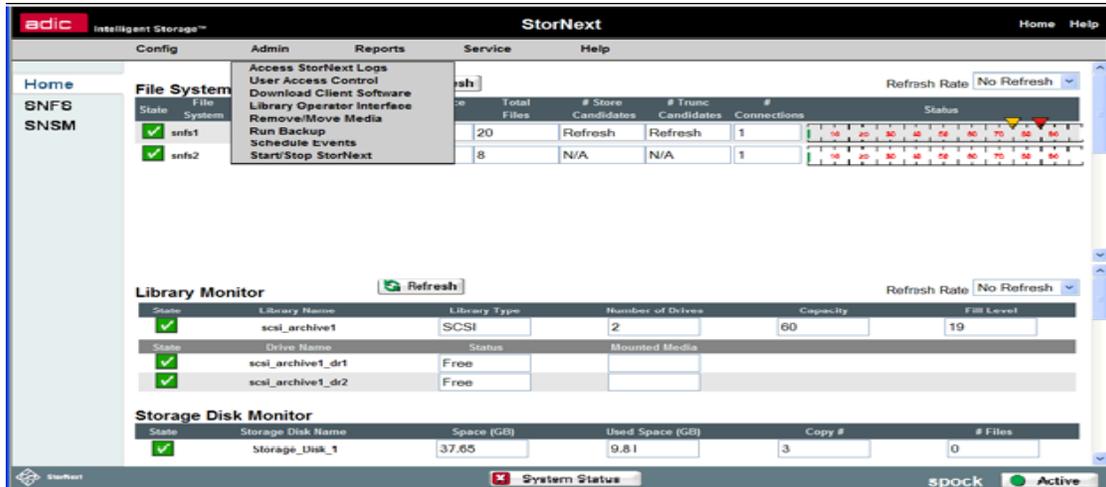
Order	Role	Task	Section	Complete?
1	Archive Manager	Executing a StorNext Backup	(P) 17.7.1	
2	Archive Manager	Scheduling a StorNext Backup	(P) 17.7.2	

17.7.1 Executing a StoreNext 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 enable you to control day-to-day operations of StorNext. The Admin menu contains these options:



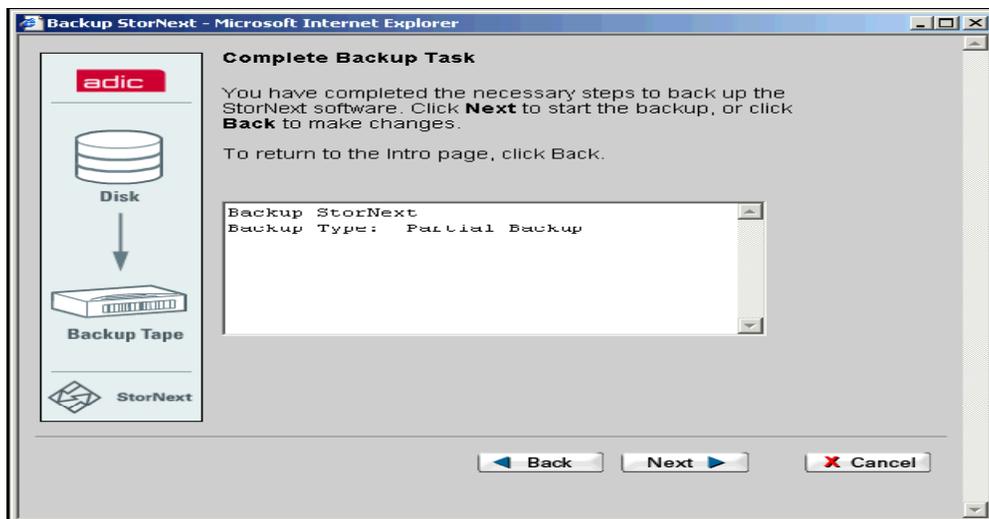
- **Access StorNext Logs:** Access logs of StorNext operations
- **User Access Control:** Control user access to StorNext tasks
- **Download Client Software:** Download SNFS client software
- **Library Operator Interface:** Enter or eject media from the Library Operator Interface
- **Remove/Move Media:** 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, and the **Backup StorNext** screen appears.



- 4 Select the type of backup you want run, **Full** or **Partial**, and click **Next**. The Complete Backup Task screen appears

| NOTE: These backups DO NOT backup user data.



- 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:

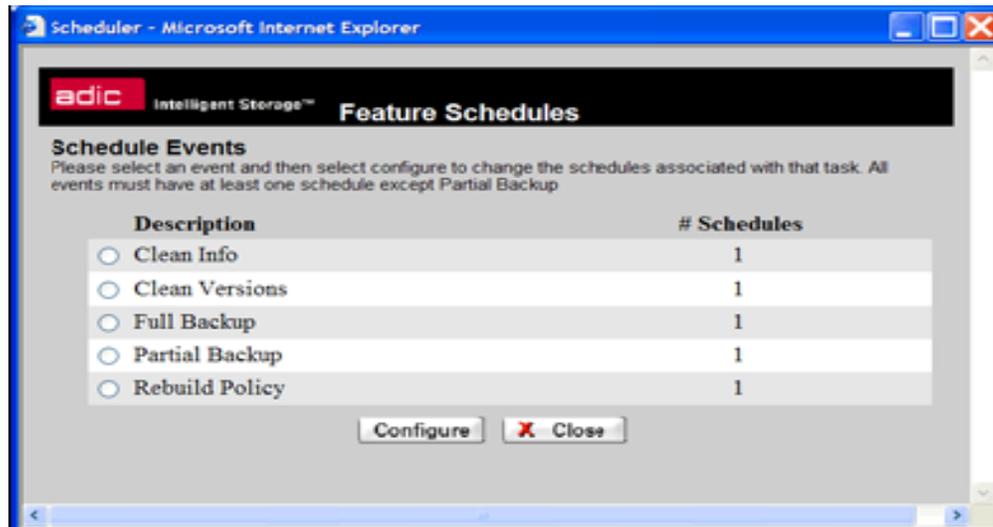
- **Clean Info:** Scheduled background operation for removing knowledge of media from StorNext
- **Clean Versions:** Clean old inactive versions of files
- **Full Backup:** By default, a full backup is run once a week to back up the entire database, configuration files, and the file system metadata dump 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.

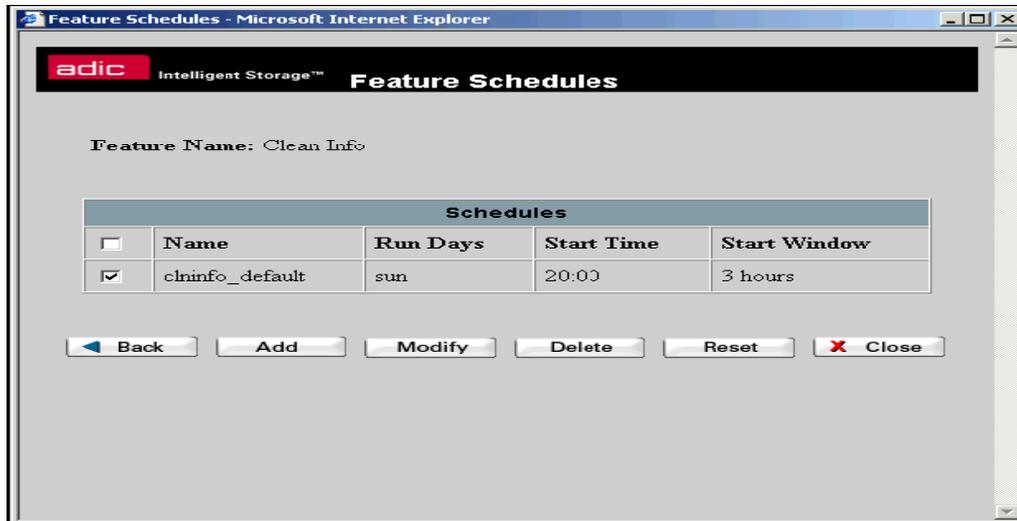
17.7.2 Scheduling a StorNext Backup

From the StorNext Home Page click **Admin > Schedule Events**.
The **Feature Schedules** screen appears.



Select a feature to schedule and click **Configure**.

The Feature Schedules screen displays the selected Feature and its current schedule.



Select a schedule and 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
-

17.8 ACSLS Procedures

For the StorageTek Powderhorn, direct control of the tape storage and handling operations is managed by the *Automated Cartridge System Library Software (ACSL)*. Full guidance for using ACSLS is provided in the *Automated Cartridge System Library Software System Administrator's Guide*. Table 17.8-1 lists the commands covered in that *Guide*.

Table 17.8-1. ACSLS Command Reference

Command	Function
Audit	Creates or updates the database inventory of the volumes in a library component.
Cancel	Cancel a current or pending request.
clear lock	Removes all active and pending locks on transports or volumes.
Dismount	Dismounts a volume.
Eject	Ejects one or more volumes from the Automated Cartridge System (ACS).
Enter	Sets a Cartridge Access Port (CAP) to enter mode.
Idle	Stops ACSLS from processing new requests.
Lock	Locks (dedicates) a volume or transport to a user.
Logoff	Exits the command processor.
Mount	Mounts a data or scratch volume.
Query	Displays the status of a library component.
Set	Sets various attributes of different library components.
Show	Displays your lock ID or user ID.
Start	Starts ACSLS request processing.
Unlock	Removes active locks on volumes or transports.
Vary	Changes the state of an ACS, LSM, CAP, transport, or port.
Venter	Enters one or more volumes with missing or unreadable labels into the ACS.

ACSL commands use the following general syntax:

command type_identifier state [options]

where **type_identifier** is the ACS component and its identifier (these are listed in the *System Administrator's Guide*), **state** is a device state for the **vary** command only, and **options** are command options (these are specified for each command in the *System Administrator's Guide*).

The two most useful commands in ACSLS are **query** and **vary**. Other frequently used commands are **enter** and **eject**, for inserting and removing cartridges, respectively. ACSLS does not have an online help facility, but if you enter a command (e.g., **vary**), it will prompt you for the parameters.

There are also several utilities provided with ACSLS. These are listed with their functions in Table 17.8-2.

Table 17.8-2. ACSLS Utilities

Utility	Function
bdb.acsss	Backs up the ACSLS database.
kill.acsss	Terminates ACSLS.
rc.acsss	Starts and recovers ACSLS.
rdb.acsss	Restores the ACSLS database.
Volrpt	Creates a volume report.
db_command	Starts or stops the ACSLS database.

To control and interact with ACSLS, you use the following user IDs:

- **acssa** lets you enter ACSLS commands from a command processor window.
- **acsss** lets you run ACSLS utilities from the UNIX command line prompt.

It is typical to log in as both user IDs to permit entering both ACSLS utilities and commands. You can, however, open a command processor window from the **acsss** user ID if you prefer to work from a single user ID. The *System Administrator's Guide* provides full details.

Table 17.8-3 provides an Activity Checklist for major ACSLS procedures addressed in this section.

Table 17.8-3. ACSLS Procedures - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Manager	Entering the Archive after StorNext is Started	(P) 17.8.1.1	
2	Archive Manager	Backing up the ACSLS Database	(P) 17.8.2.1	
3	Archive Manager	Restoring the ACSLS Database	(P) 17.8.3.1	
4	Archive Manager	Checking Cleaning Cartridges	(P) 17.8.4.1	

17.8.1 Entering the Archive After StorNext is Started

There are circumstances in which it may be necessary to enter the archive after StorNext is started. For example, there may be a requirement for maintenance that necessitates access to the robot or other area inside the Powderhorn. Another example is that it may sometime be desirable to bypass the Cartridge Access Port (CAP) when inserting tape cartridges, if there is a need to perform bulk loading of a large number of tapes, although usually this would be limited to initial loading of the volumes.

17.8.1.1 Entering the Archive After StorNext is Started

- 1 At the host for ACSLS (e.g., **e0drs03**, **g0drs03**, **l0drs02**, **n0drs03**), log in using the **acssa** user ID and password.
 - The **acssa** command-process window is displayed with the **ACSSA>** prompt.

- 2 Type **vary lsm 0,0 offline** and then press the **Return/Enter** key.
 - The access port is unlocked (audible unlatching sound).
- 3 Use the key to unlatch and open the access door.
 - A red **DO NOT ENTER** warning is visible inside the enclosure.

Warning

If it is necessary to enter the STK Powderhorn after StorNext is started, it is necessary to perform the following step to avoid hazard and ensure safety of personnel and equipment.

- 4 Remove the key from the door to ensure that no one inadvertently locks the enclosure with someone inside.
 - The red **DO NOT ENTER** warning is extinguished and a green **ENTER** message is displayed inside the enclosure.
 - 5 Upon leaving the enclosed area, insert the key in the access door and latch the door.
 - The LED display indicates that the door is locked.
 - 6 At the ACSLS host, type **vary lsm 0,0 online** and then press the **Return/Enter** key. After a few seconds, the archive robots execute an initialization sequence and the LSM is back online.
-

17.8.2 Backing Up the ACSLS Database

ACSLS provides the **bdb.acsss** utility to back up the database. It is advisable to run this utility when there has been a change in the archive volume structure (e.g., upon addition or removal of volumes). In the event of database loss, it is possible to re-create the database even if there is no backup available, by using the **audit** command to inventory the archive. However, for a large storage facility, creating the database this way may take several hours. If there is a backup available, the database can be restored easily and quickly (refer to Section 17.8.3)

17.8.2.1 Backing Up the ACSLS Database

- 1 At the host for ACSLS (e.g., **e0drs03**, **g0drs03**, **l0drs02**, **n0drs03**), log in using the **acsss** user ID and password.
 - The **acsss** command-process window is displayed with the **ACSSS>** prompt.
- 2 Ensure that there is a tape in the backup drive (device **dev/rmt/0**), a streaming tape drive attached to each ACSLS workstation.

- 3 Type **bdb.acsss**, and then press the **Return/Enter** key.
 - If you enter **bdb.acsss** with no options, the backup utility defaults to the default tape device attached and configured to the ACSLS server.
 - The system displays the following message.
Check tape device (/dev/rmt/0) to make sure you have a tape in the tape drive.
[Hit RETURN to continue or Ctrl-C to exit]
 - 4 Press the **Return/Enter** key.
 - The **bdb.acsss** utility backs up the ACSLS database and miscellaneous library resource files.
-

17.8.3 Restoring the ACSLS Database

ACSLs provides the **rdb.acsss** utility to restore the database in case of severe disk or data problems. If you have made regular backups, it should be possible to restore the database with little or no loss of data. Restoring the database is likely to be necessary if there has been a system crash, or if the database cannot be started or has a physical or logical error.

17.8.3.1 Restoring the ACSLS Database

- 1 At the host for ACSLS (e.g., **e0drs03**, **g0drs03**, **l0drs02**, **n0drs03**), log in using the **acsss** user ID and password.
 - The **acsss** command-process window is displayed with the **ACSSS>** prompt.
 - 2 Load the restore tape into the backup drive.
 - 3 Type **rdb.acsss**, and then press the **Return/Enter** key.
 - If you enter **bdb.acsss** with no options, the backup utility defaults to the default tape device attached and configured to the ACSLS server.
 - The system displays the following message.
Check tape device (/dev/rmt/0) to make sure you have a tape in the tape drive.
[Hit RETURN to continue or Ctrl-C to exit]
 - 4 Press the **Return/Enter** key.
 - The **rdb.acsss** utility restores the ACSLS database and miscellaneous library resource files.
-

17.8.4 Checking Cleaning Cartridges

The Automated Cartridge System Library Software (ACSLs) schedules and implements routine cleaning of the system tape drives after a set usage time tracked by the software, using cleaning volumes from a cleaning volume group designated for that purpose. The ACSLS software also tracks the number of times a cleaning tape is used, and will not use a cleaning tape that has been used the maximum set number of times (usually set at 100 for the 9940 drives). It is the

responsibility of the Archive Manager to monitor cleaning tape usage periodically, to ensure that usable cleaning tapes remain available to the system.

17.8.4.1 Checking Cleaning Cartridges

1 At the host for ACSLS (e.g., **e0drs03**, **g0drs03**, **l0drs02**, **n0drs03**), log in using the **acssa** user ID and password.

- The **acssa** command-process window is displayed with the **ACSSA>** prompt.

2 Type **query clean all**, and press the **Return/Enter** key.

- **Note:** The command may be abbreviated to **qu cl a**.
- ACSLS displays information on the status of the cleaning volumes in format similar to the following:

```
2001-10-04 08:50:54 Cleaning Cartridge Status
Identifier Home Location Max Usage Current Usage Status Type
9840C1 0, 0, 3, 2, 2 100 38 home STK1U
9840C2 0, 0,13, 1, 3 100 0 home STK1U
9940C1 0, 0, 1, 4,19 100 7 home STK2W
```

Note: If it is desirable or necessary to change the maximum number of uses permitted for a cleaning volume, the change can be accomplished with the command **set clean <max_usage> <vol_id>** where **max_usage** (e.g. 100) is the maximum number of uses for that volume and **vol_id** is the volume id of that cleaning cartridge.

17.9 Data Pool Maintenance Tasks

17.9.1 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 in a cron job, but may be invoked manually. 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 17.9-1 provides an Activity Checklist for Data Pool Maintenance Tasks addressed in this section.

Table 17.9-1. Data Pool Maintenance Tasks - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Technician	Launch the DPM GUI	(P) 17.9.1.1	
2	Archive Technician	Shut Down the DPM GUI	(P) 17.9.1.2	
3	Archive Technician	Monitor Data Pool Active Insert Processes	(P) 17.9.1.3	
4	Archive Technician	View a List of Data Pool File Systems	(P) 17.9.2.1	
5	Archive Technician	Add a Data Pool File System	(P) 17.9.2.2	
6	Archive Technician	Modify a Data Pool File System	(P) 17.9.2.3	
7	Archive Technician	View cloud Cover Information	(P) 17.9.3.1	
8	Archive Technician	Add New Cloud Cover Information	(P) 17.9.3.2	
9	Archive Technician	Modify Cloud Cover Information	(P) 17.9.3.3	
10	Archive Technician	Delete Cloud Cover Information	(P) 17.9.3.4	
11	Archive Technician	Check the Status of Batch Inserts	(P) 17.9.4.1	
12	Archive Technician	Checking the Data Pool Insert Queue and Cancel a Data Pool Insert Action	(P) 17.9.5.1	
13	Archive Technician	View DPM Configuration Parameter Values	(P) 17.9.6.1	
14	Archive Technician	Modify DPM Configuration Parameter Values	(P) 17.9.6.2	
15	Archive Technician	View DPM Aging Parameter Values	(P) 17.9.7.1	
16	Archive Technician	Modify DPM Aging Parameter Values	(P) 17.9.7.2	
17	Archive Technician	View Collection Group and Collection Information	(P) 17.9.8.1	
18	Archive Technician	Modify Collection Groups	(P) 17.9.8.2	
19	Archive Technician	Add a Collection Group	(P) 17.9.8.3	
20	Archive Technician	Add an ECS Collection Group	(P) 17.9.8.4	
21	Archive Technician	Modify an ECS Collection Group	(P) 17.9.8.5	
22	Archive Technician	View a List of Themes	(P) 17.9.9.1	
23	Archive Technician	Filter a List of Themes	(P) 17.9.9.2	
24	Archive Technician	Modify a Theme	(P) 17.9.9.3	
25	Archive Technician	Add a Theme	(P) 17.9.9.4	
26	Archive Technician	Delete a Theme	(P) 17.9.9.5	

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 17.9-1).

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.

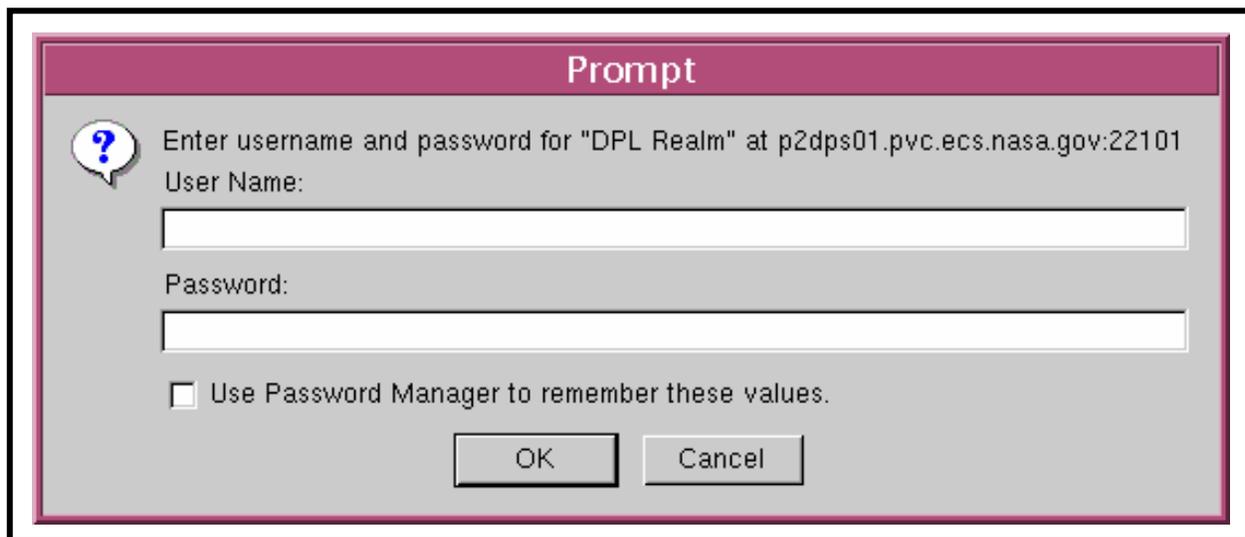


Figure 17.9-1. Security Login Prompt

Figure 17.9-2 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.

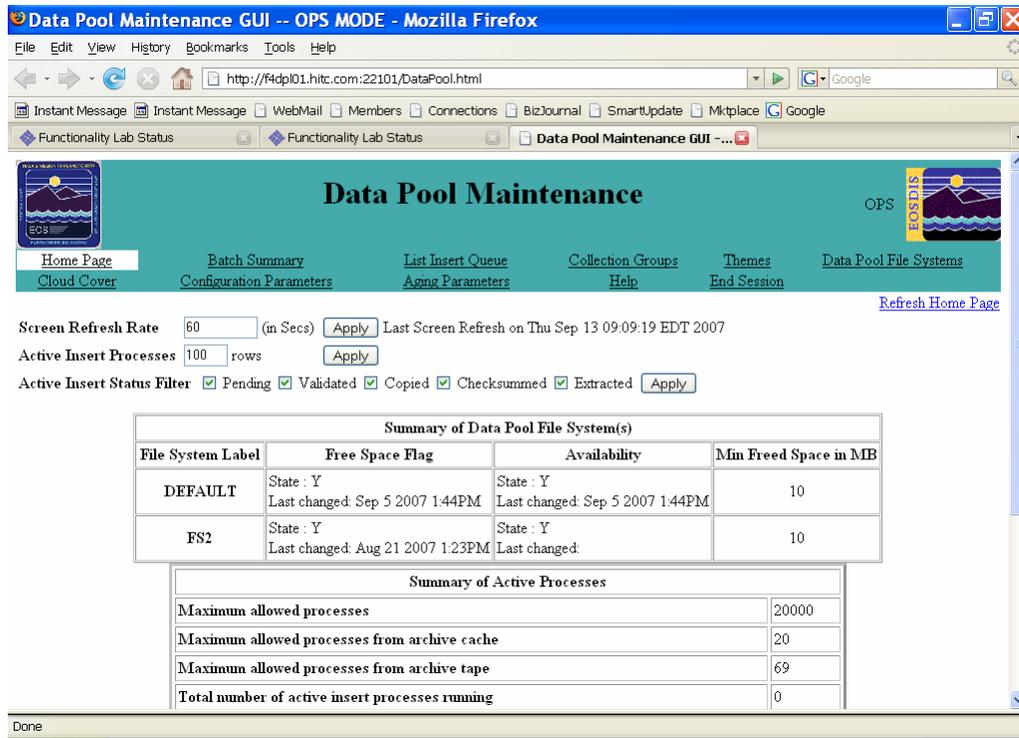


Figure 17.9-2. DPM GUI Home Page

The **DPM GUI Home Page** (Figure 17.9-2) 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.

17.9.1.1 Launch the DPM GUI

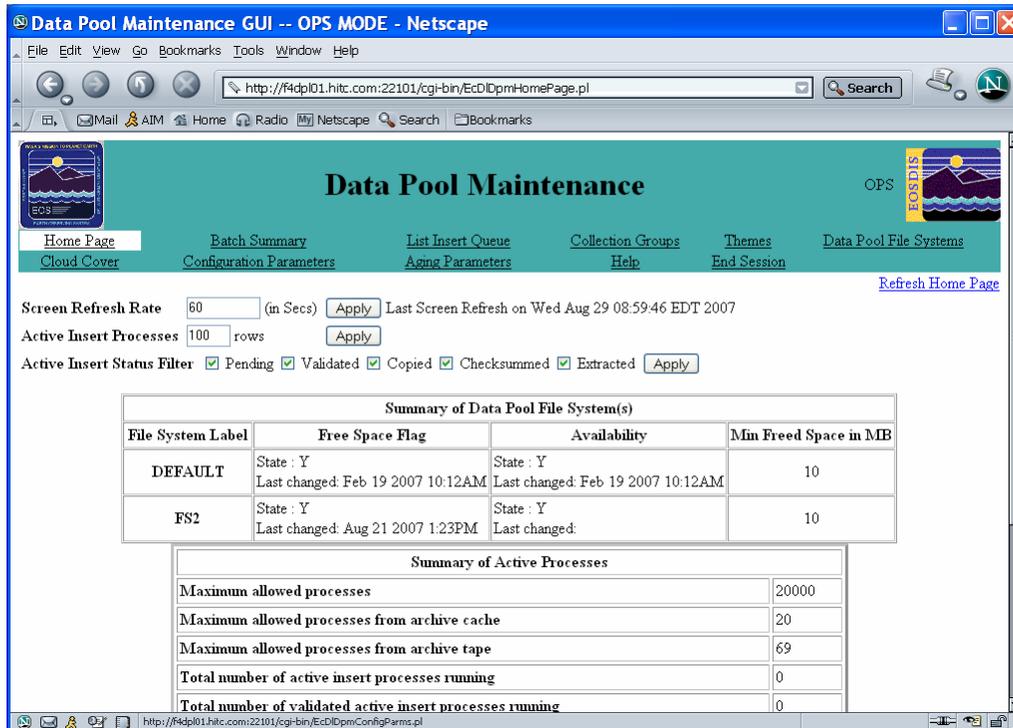
- 1 At the UNIX command shell prompt, type **setenv DISPLAY *clientname*:0.0** and then press the **Return/Enter** key.
 - For *clientname*, use either the local terminal/workstation IP address or its machine name.
- 2 Start the log-in to a Firefox host by typing **/tools/bin/ssh *hostname* (x4dpl01)** 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's.
 - 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.



- **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.

17.9.1.2 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.

17.9.1.3 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 Label** (label representing an existing Data Pool file system).
 - **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 cache.**
 - **Maximum allowed processes from 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 cache.**
 - **Number of active insert processes using 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.

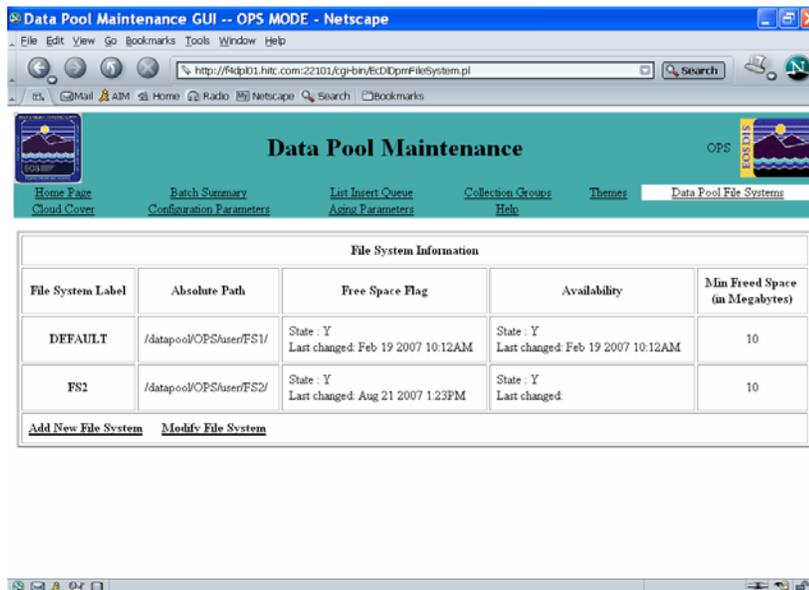
- 3** To obtain an immediate screen refresh, click on the **Refresh Home Page** link near the upper right corner of the display.
- The displayed data are updated.

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.
-

17.9.2 Data Pool File Systems

Figure 17.9-3 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).



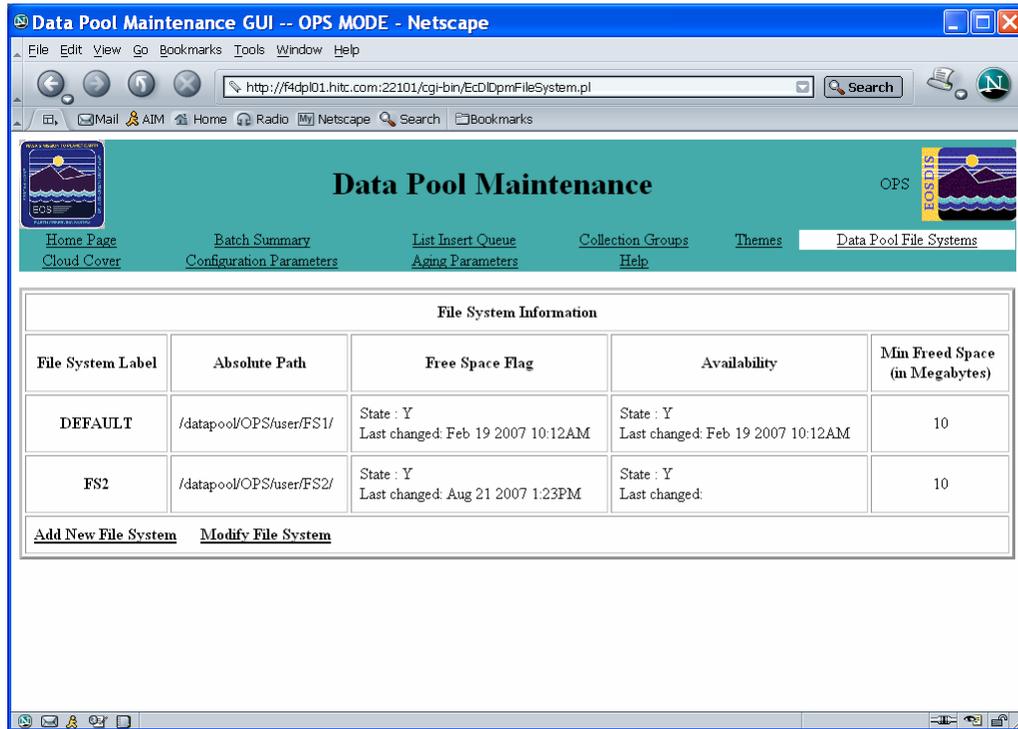
File System Information				
File System Label	Absolute Path	Free Space Flag	Availability	Min Freed Space (in Megabytes)
DEFAULT	/datapool/OPSUser/FS1/	State : Y Last changed: Feb 19 2007 10:12AM	State : Y Last changed: Feb 19 2007 10:12AM	10
FS2	/datapool/OPSUser/FS2/	State : Y Last changed: Aug 21 2007 1:23PM	State : Y Last changed:	10

[Add New File System](#) [Modify File System](#)

Figure 17.9-3. Data Pool File System Information Page

17.9.2.1 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.



- 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 Label.**
 - **Absolute Path.**
 - **Free Space Flag.**
 - **Availability.**
 - **Min Free Space (in Megabytes).**
 - The following links are available on the **File System Information** page:
 - **Add New File System.**
 - **Modify File System.**
-

Clicking on the **Add New File System** link takes the full-capability operator to the **Add New File System** page shown in Figure 17.9-4. The operator needs to add data in the following five fields:

- 1) [File System] **Label**: a label representing an existing Data Pool file system;
- 2) **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);
- 3) **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);
- 4) **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);
- 5) **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.

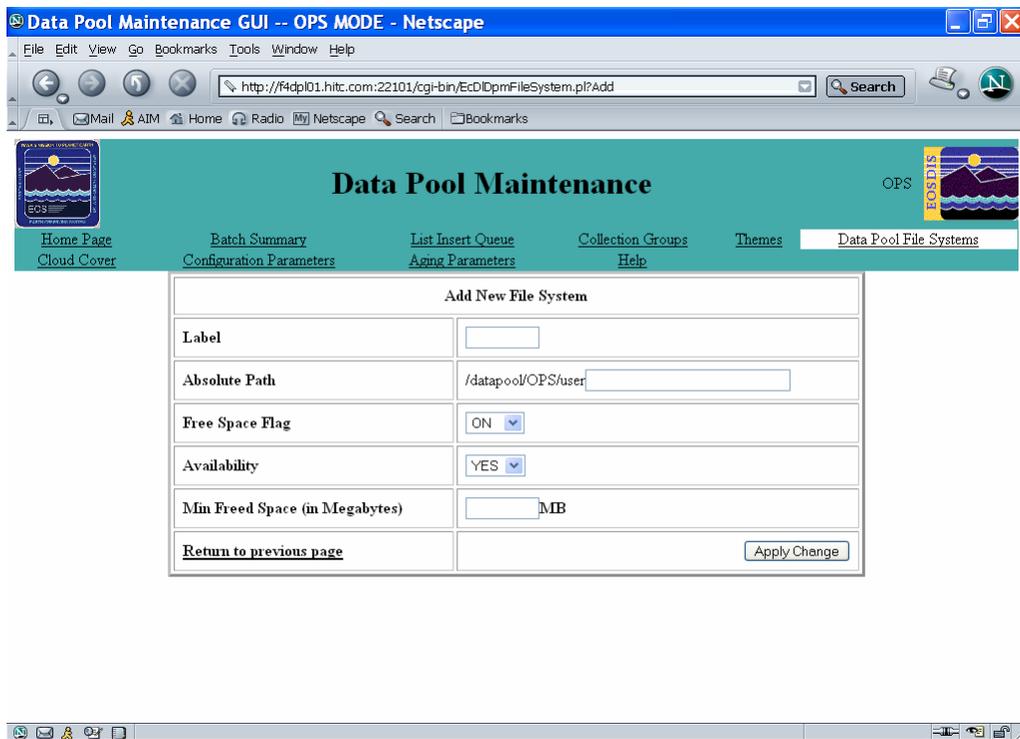
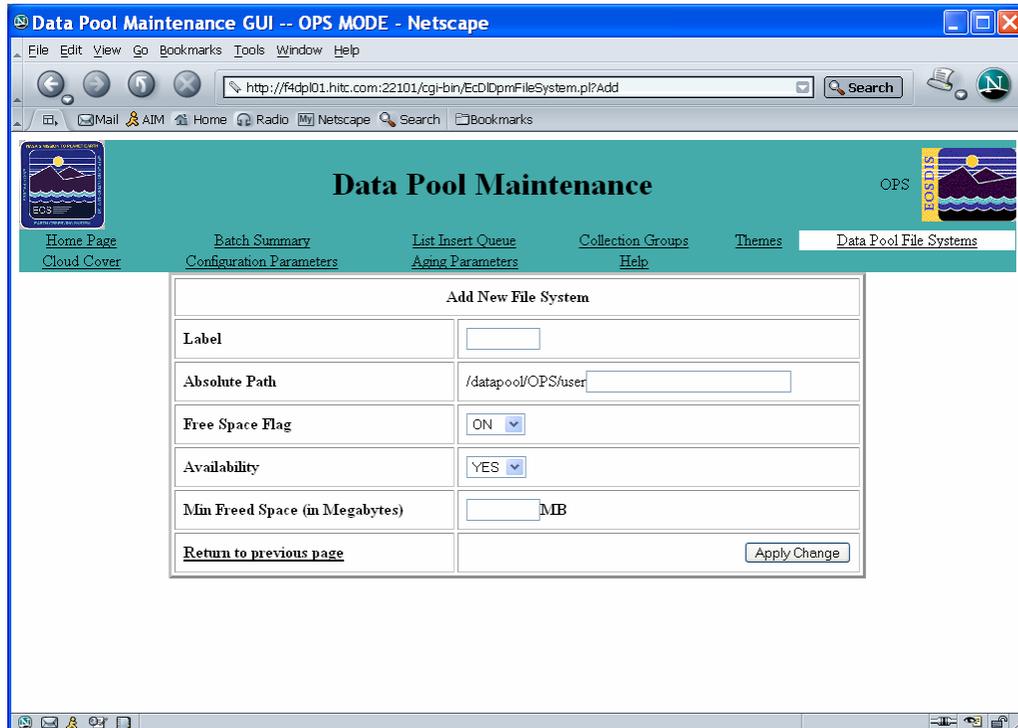


Figure 17.9-4. Add New File System Page

17.9.2.2 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.



The screenshot shows a Netscape browser window titled "Data Pool Maintenance GUI -- OPS MODE - Netscape". The address bar shows the URL "http://fdp101.hitc.com:22101/cgi-bin/EcDIDpmFileSystem.pl?Add". The page has a teal header with the title "Data Pool Maintenance" and a navigation menu with links: Home Page, Batch Summary, List Insert Queue, Collection Groups, Themes, and Data Pool File Systems. Below the header is a form titled "Add New File System" with the following fields and options:

Add New File System	
Label	<input type="text"/>
Absolute Path	/datapool/OPS/user <input type="text"/>
Free Space Flag	<input type="button" value="ON"/> <input type="button" value="OFF"/>
Availability	<input type="button" value="YES"/> <input type="button" value="NO"/>
Min Freed Space (in Megabytes)	<input type="text"/> MB
Return to previous page	<input type="button" value="Apply Change"/>

- 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 select a free space flag option click on the appropriate choice from the option list.
 - **ON** should be selected if there is enough free space in the file system for inserts.
 - **OFF** should be selected if there is not enough free space in the file system for inserts.

- 8 To display availability options click on the **Availability** option button.
 - **Availability** options are displayed (i.e., **YES** and **NO**).
 - 9 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.
 - 10 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.
 - 11 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 17.9-5. 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.

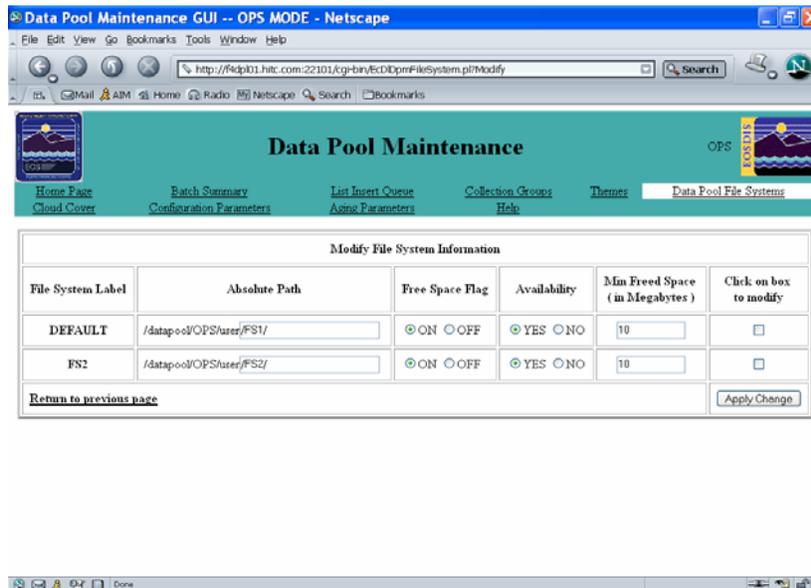
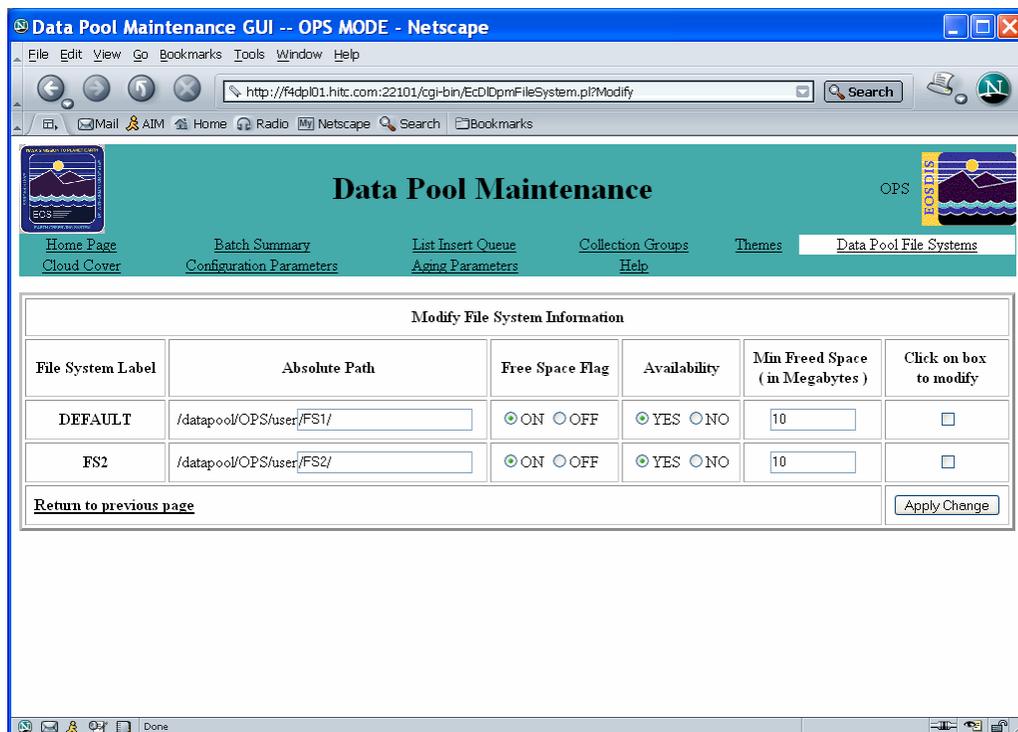


Figure 17.9-5. Modify File System Information Page

17.9.2.3 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.



- 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.
-

17.9.3 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 17.9-2. The link takes the operator to the **Cloud Cover Information** page shown in Figure 17.9-6. The page displays the information concerning the sources of cloud cover; i.e., the **Source Type**, **Source Name**, and **Source Description**.

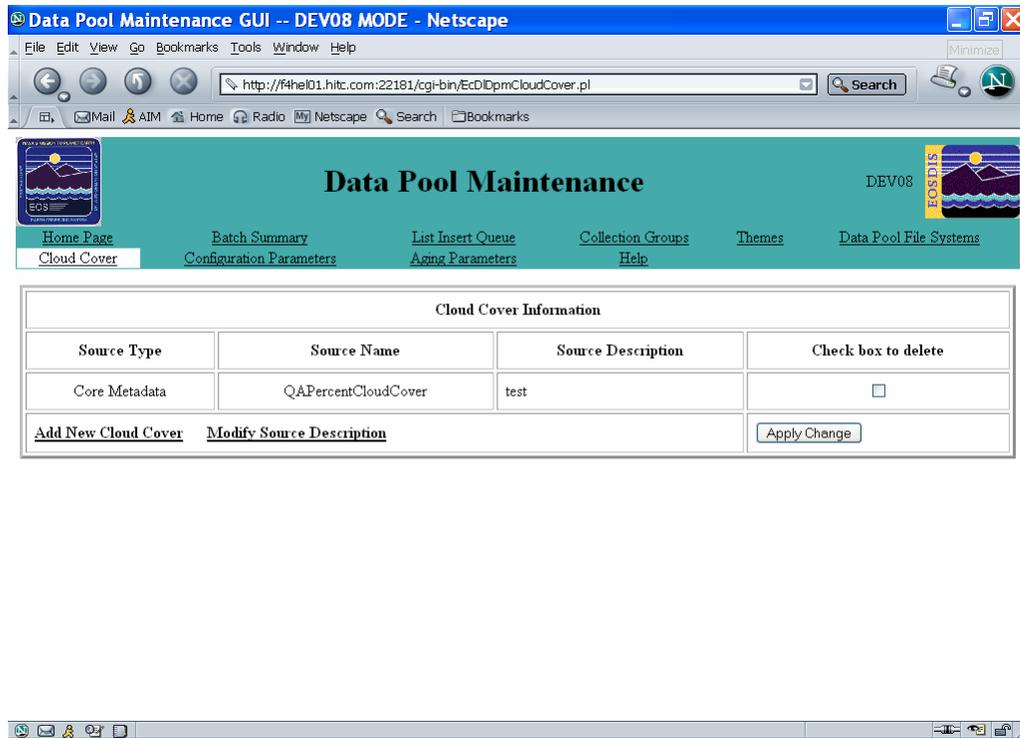
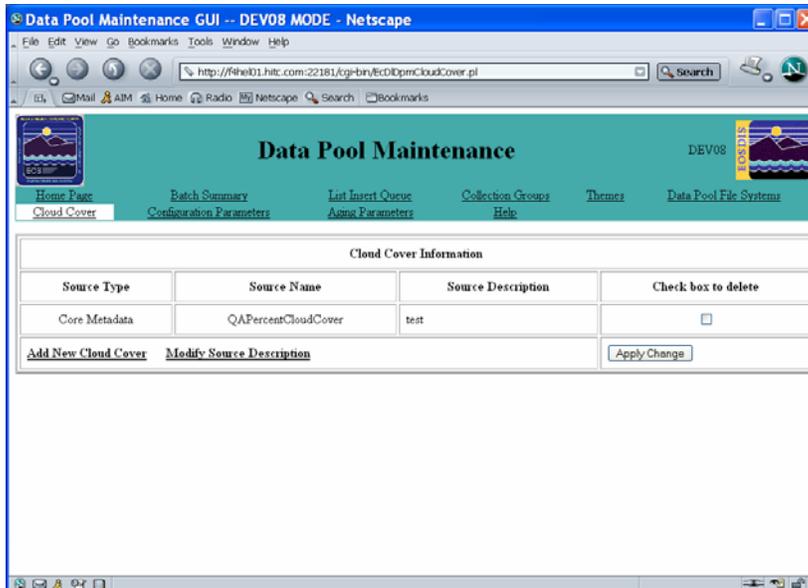


Figure 17.9-6. Cloud Cover Information Page

17.9.3.1 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.



- 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 17.9-6. The link takes the operator to the **Add New Cloud Cover Information** page shown in Figure 17.9-7. 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 Science Data Server database. The new cloud cover source is added to the Data Pool database and the **Cloud Cover Information** page is refreshed.

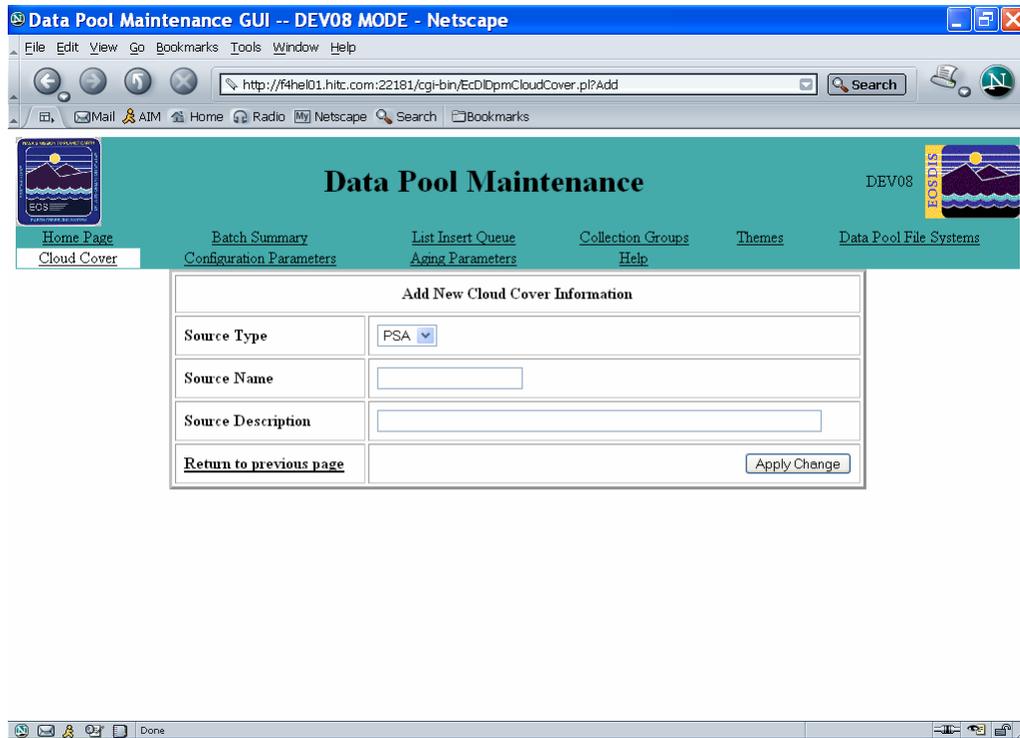
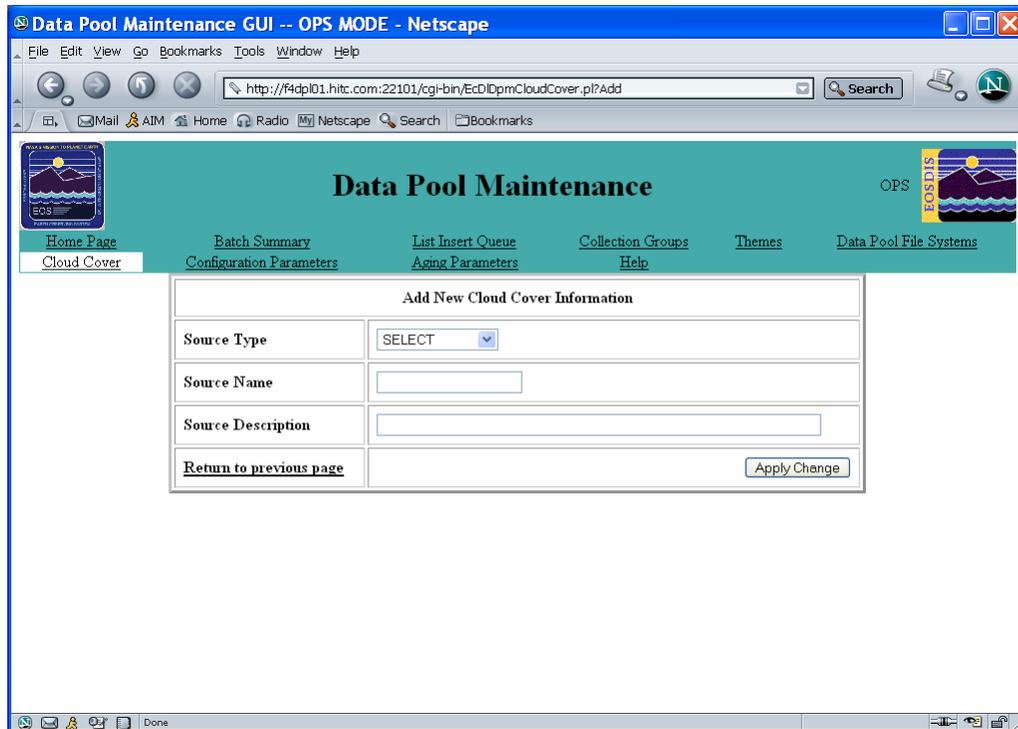


Figure 17.9-7. Add New Cloud Cover Information Page

17.9.3.2 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.



- 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 Science Data Server 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 17.9-6. The link takes the operator to the **Modify Source Description** page shown in Figure 17.9-8. 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.

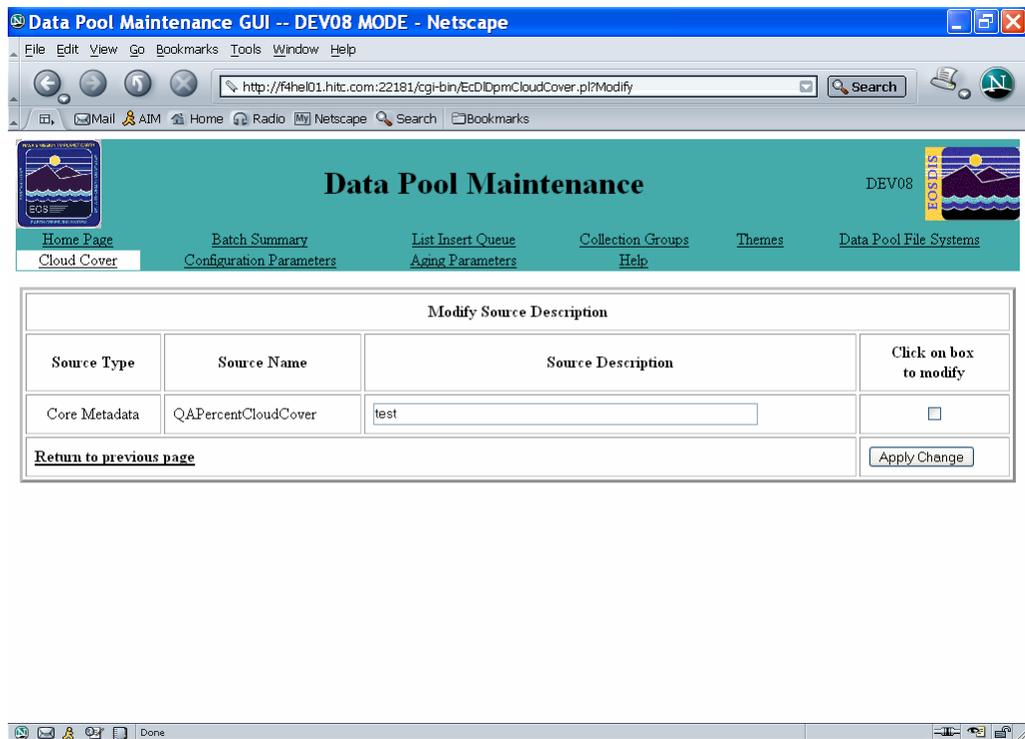
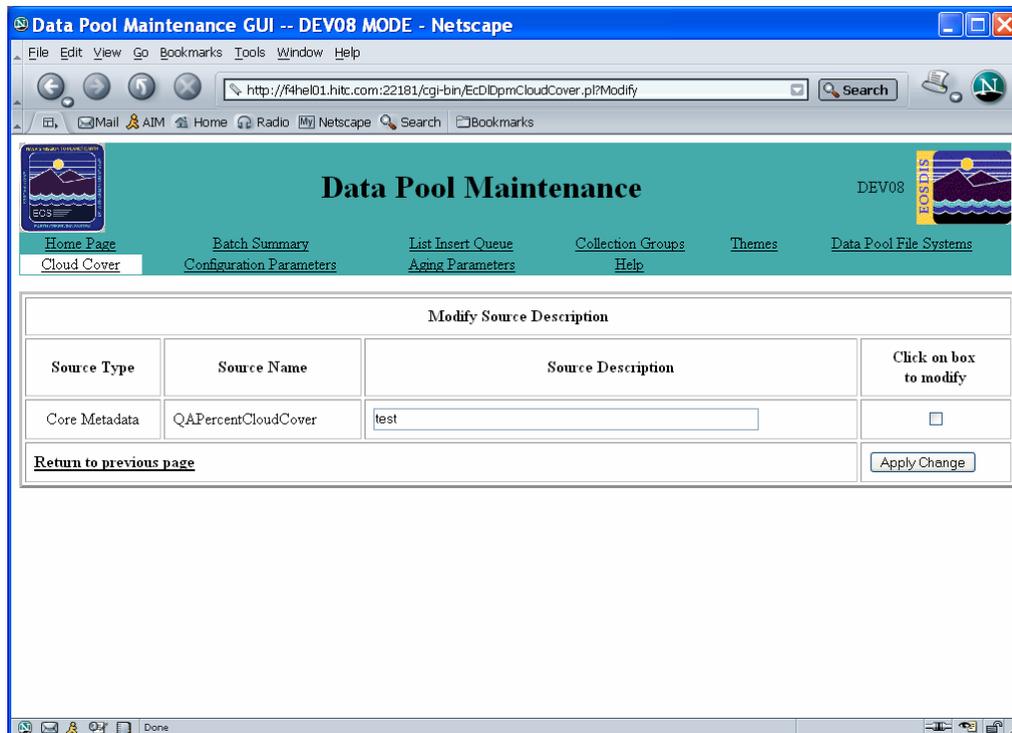


Figure 17.9-8. Modify Source Description Page

17.9.3.3 Modify Cloud Cover Source Descriptions

- 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.



- 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.

17.9.3.4 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.
-

17.9.4 Batch Summary

Figure 17.9-9 illustrates the **Batch Summary** page, which is accessible from the **Batch Summary** link on the **DPM GUI Home Page** (Figure 7.9-2). 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.

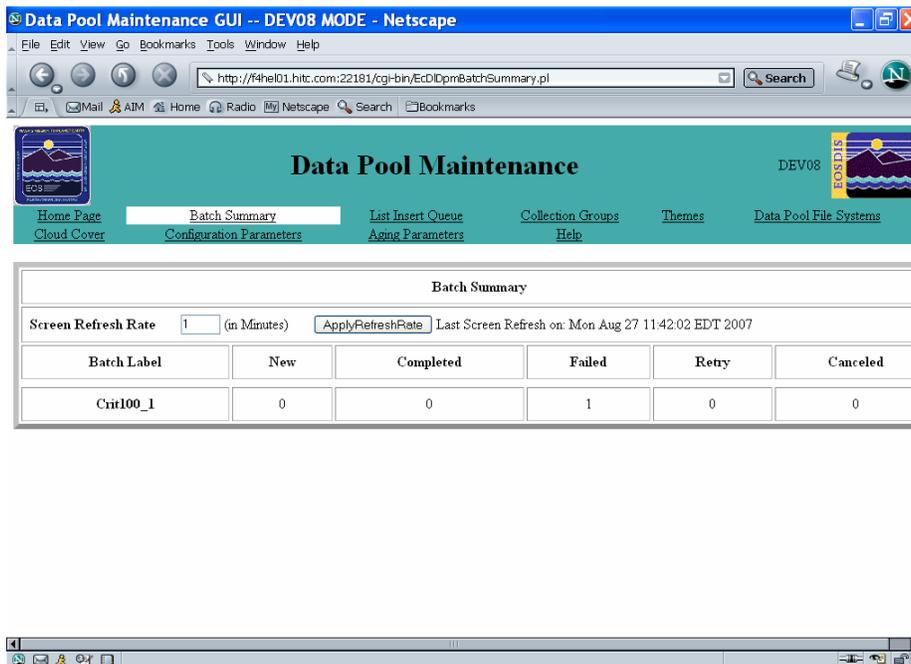
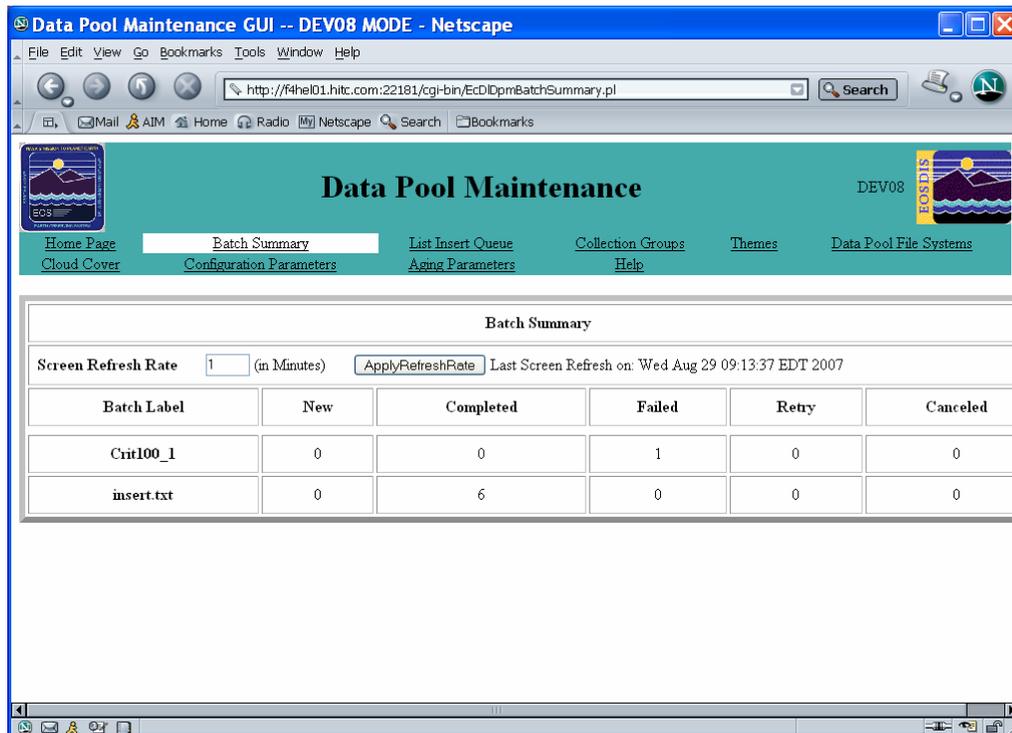


Figure 17.9-9. Batch Summary Page

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

17.9.4.1 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.



- 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.
-

17.9.5 List Insert Queue

Figure 17.9-10 illustrates the **List Insert Queue** page, which is accessible from the **List Insert Queue** link on the **DPM GUI Home Page** (Figure 17.9-2). 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.

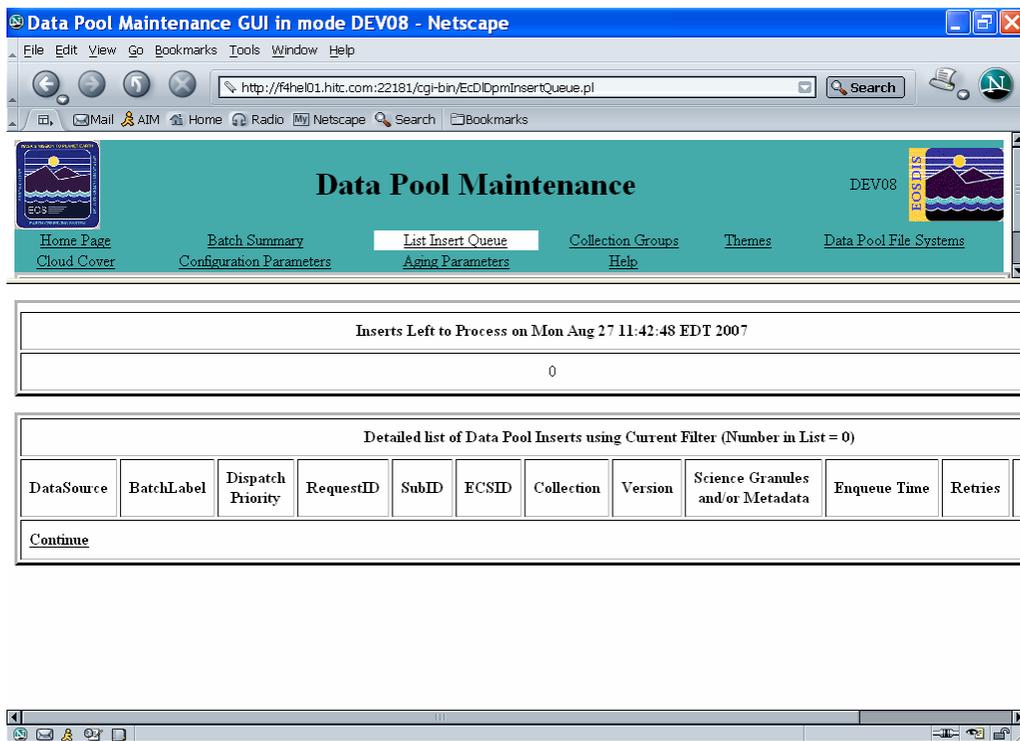


Figure 17.9-10. List Insert Queue Page

The batch insert utility can be used to insert non-ECS data into the Data Pool. On the **DPM GUI List Insert Queue** page, non-ECS data insert actions are identified by the entry “NONECS” in the **Data Source** column. XML file and path name for a non-ECS granule insert action can be viewed by clicking on “NONECS” in the **Data Source** column. Figure 17.9-11 shows the appearance of the pages displaying absolute .XML file path and file content.

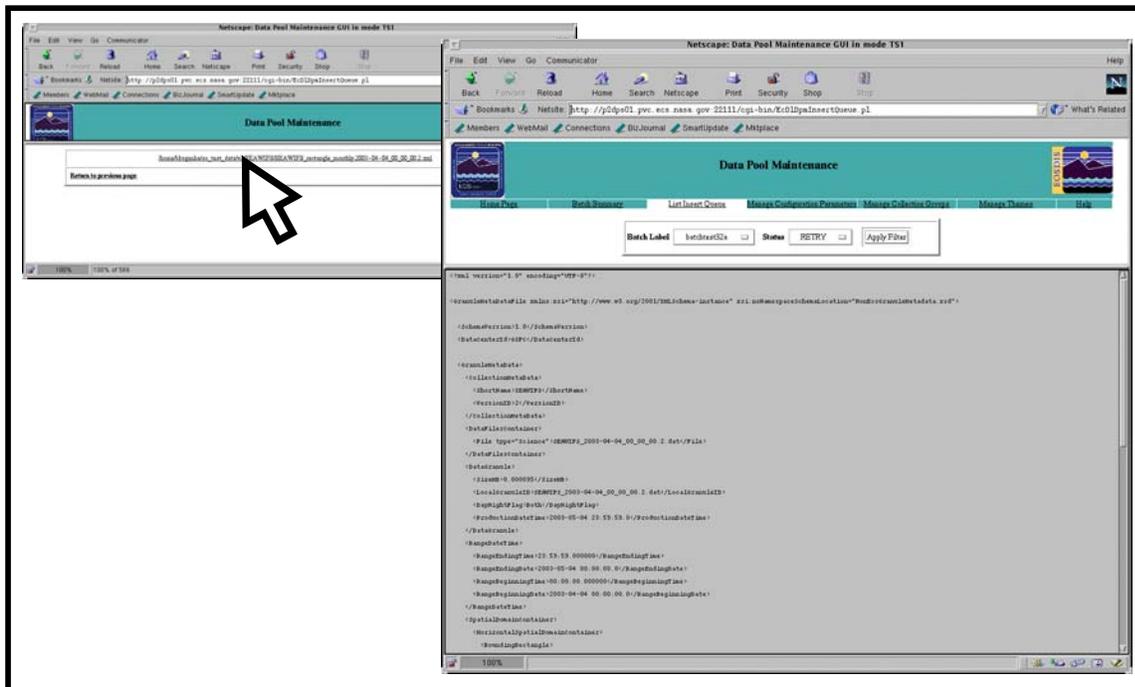
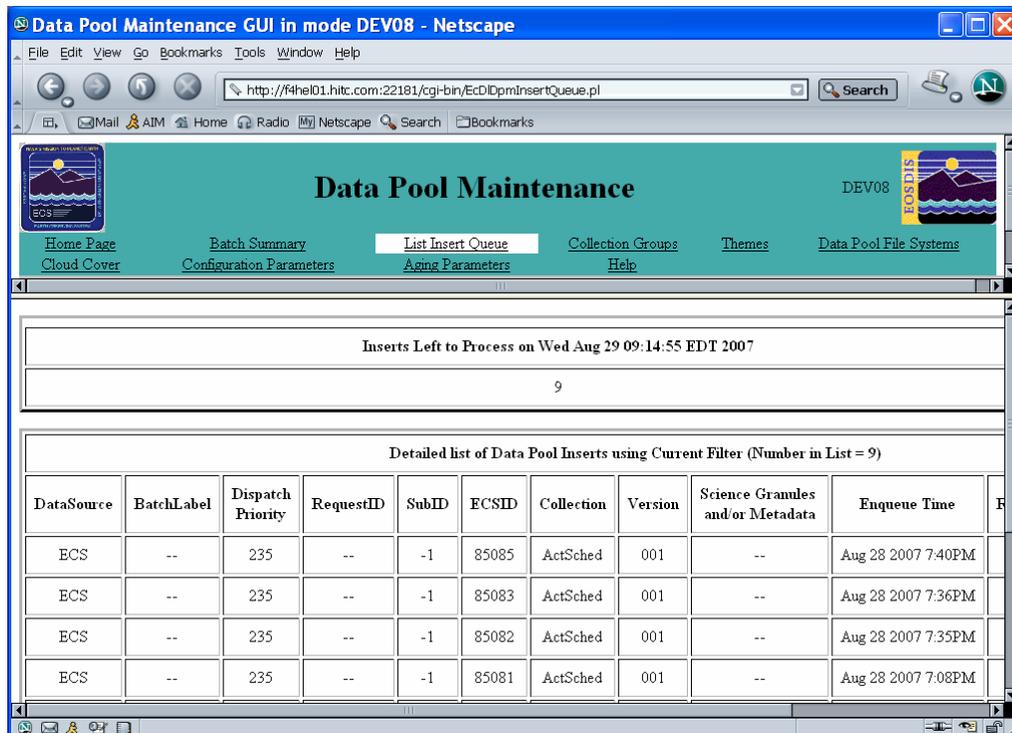


Figure 17.9-11. Sample DPM GUI Pages for .XML File Path and Content

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.

17.9.5.1 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.



- 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.
-

17.9.6 Configuration Parameters

Figure 17.9-12 shows the **List of Configuration Parameters** page, which is accessible from the **Configuration Parameters** link on the **DPM GUI Home Page** (Figure 17.9-2). 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).

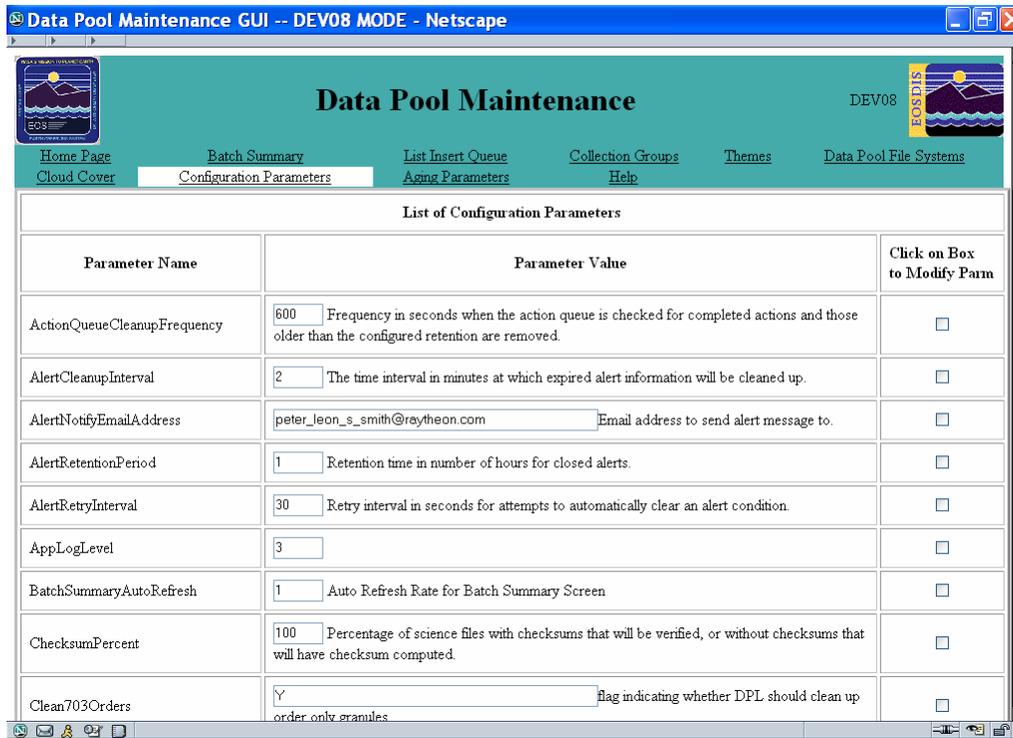


Figure 17.9-12. List of Configuration Parameters Page

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

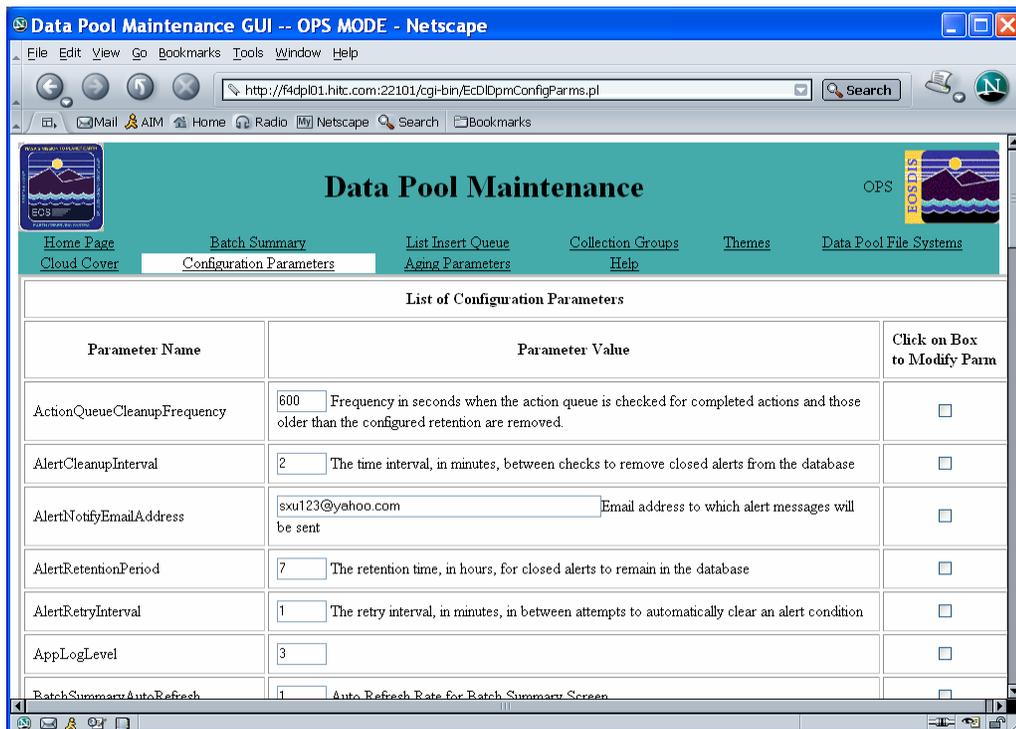
- ActionQueueCleanupFrequency - frequency in seconds when the action queue is checked for completed actions and those older than the configured retention period are removed.
- BatchSummaryAutoRefresh – autorefresh rate for the Batch Summary page.
- DefaultRetentionPeriod - default retention period in 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.

- HEGCleanupAge – HDF-EOS to GeoTIF Converter (HEG) cleanup age in days.
- IdleSleep - number of seconds to sleep when there is nothing to do.
- 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.
- InsertRetryWait - number of seconds to wait before an insert that failed should be resubmitted (if it can be retried).
- MAX_READ_DRIVES_x0xxgmn – (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.
- 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.
- NumOfAllowedInsertProcesses - maximum number of insert processes running at any time.
- NumOfAllowedNonCacheProcesses - maximum number of insert processes that require access to tape.
- 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.
- 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.

17.9.6.1 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.



- 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.
 - AppLogLevel.
 - BatchSummaryAutoRefresh.
 - ChecksumPercent.
 - Clean703Orders.
 - DatabaseRetryCount.
 - DatabaseRetryInterval.
 - DefaultRetentionPeriod.
 - DefaultRetentionPriority.
 - DeleteCompletedActionsAfter.
 - DisplayAIPChunkSize.
 - FileSystemCheckInterval.
 - FilterChecksumAIP.
 - FilterCopiedAIP.
 - FilterExtractedAIP.
 - FilterPendingAIP.
 - FilterValidAIP.
 - FreeSpaceResumePercent.
 - GranuleLockRetentionPeriod.
 - GranuleOMLockRetentionPeriod.
 - HEGCleanupAge.
 - IdleSleep.

- InCacheTimeLimit.
 - InsertRetryWait.
 - MAX_READ_DRIVES_x0xxgmn.
 - MFSONInsert.
 - MaxConcurrentBandExtract.
 - MaxConcurrentDPIUThreads.
 - MaxConcurrentEventThreads.
 - MaxConcurrentPublish.
 - MaxConcurrentReadsPerTape.
 - MaxConcurrentRegister.
 - MaxConcurrentValidate.
 - MaxConsecutiveErrors.
 - MaxInsertRetries.
 - MaxReadDrivesPerArchive.
 - MaxTapeMountPerRequest.
 - NewActionCheckFrequency.
 - NumOfAllowedCacheProcesses.
 - NumOfAllowedInsertProcesses.
 - NumOfAllowedNonCacheProcesses.
 - OnTapeTimeLimit.
 - OrdersOnlyFSLabel.
 - PerfLogLevel.
 - RefreshRate.
 - RunAwayCheckFrequency.
 - RunawayDuration.
 - SizeOfInsertQueueList.
 - StartUpWait.
- 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.

17.9.6.2 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.
-

17.9.7 Aging Parameters

Figure 17.9-13 shows the **List of Aging Parameters** page, which is accessible from the **Aging Parameters** link on the **DPM GUI Home Page** (Figure 17.9-2). 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.

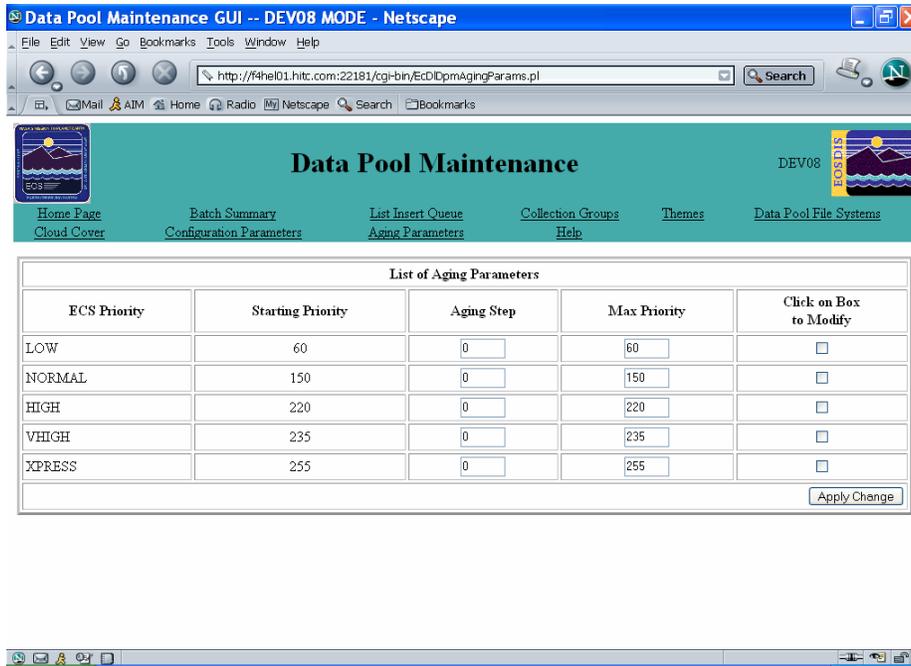
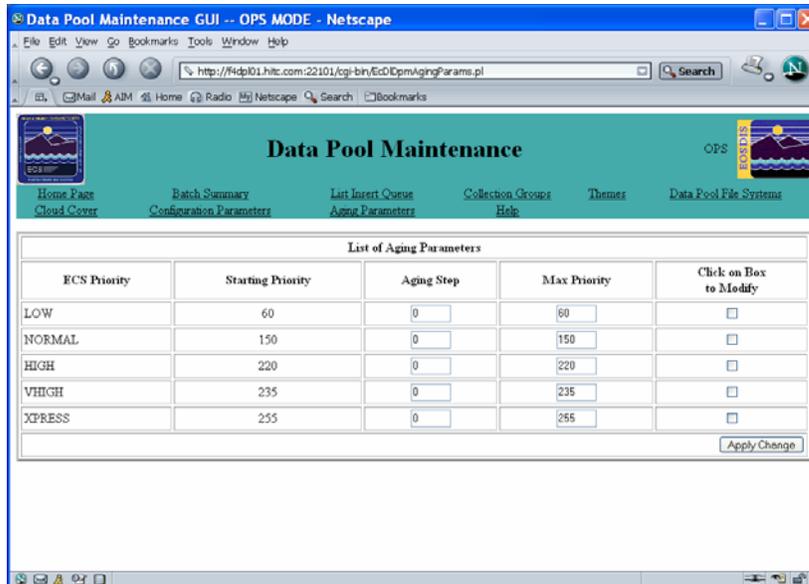


Figure 17.9-13. List of Aging Parameters Page

17.9.7.1 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.



- 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 Parm** (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.

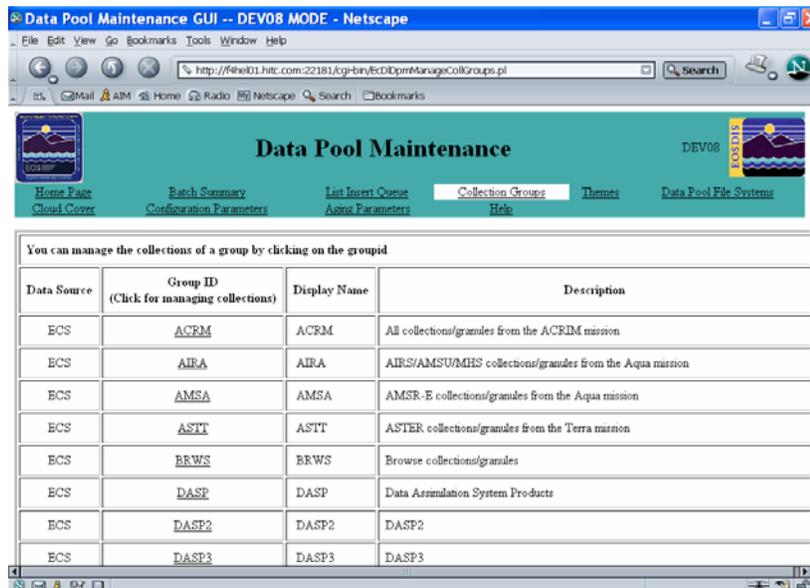
For the aging step, and maximum priority associated with each ECS priority there is a text box so the full-capability operator can assign a new value to the parameter when necessary. In addition, there is a check box that the full-capability operator uses to mark parameters with values to be modified.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

17.9.7.2 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.

17.9.8 Collection Groups

Figure 17.9-14 illustrates the Collection Groups page and is accessible from the **Collection Groups** link on the **Home Page** (Figure 17.9-2). The page lists the collection groups, providing for each the **Data Source** (ECS or NON-ECS), **Group ID**, **Display Name**, and a brief **Description** of the collection group.



Data Source	Group ID (Click for managing collections)	Display Name	Description
ECS	ACRM	ACRM	All collections/granules from the ACRM mission
ECS	AIRA	AIRA	AIRS/AMSU/MHS collections/granules from the Aqua mission
ECS	AMSA	AMSA	AMSR-E collections/granules from the Aqua mission
ECS	ASTT	ASTT	ASTER collections/granules from the Terra mission
ECS	BRWS	BRWS	Browse collections/granules
ECS	DASP	DASP	Data Assimilation System Products
ECS	DASP2	DASP2	DASP2
ECS	DASP3	DASP3	DASP3

Figure 17.9-14. 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 17.9-15 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.

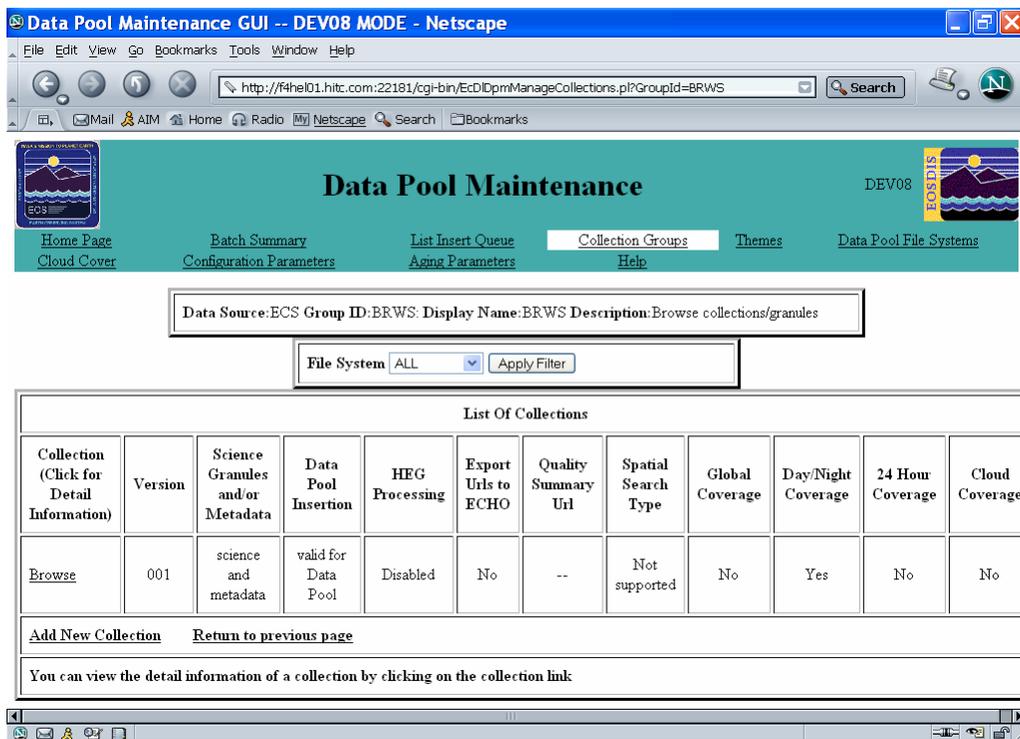


Figure 17.9-15. List of Collection Page

At the bottom of the **List of Collections** page, there is a link that permits a full-capability operator to **Add New Collection** to the collection group. A click on one of the **Collection (Click for Detail Information)** ID links brings up a collection **Detail Information** page (Figure 17.9-16) 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**.

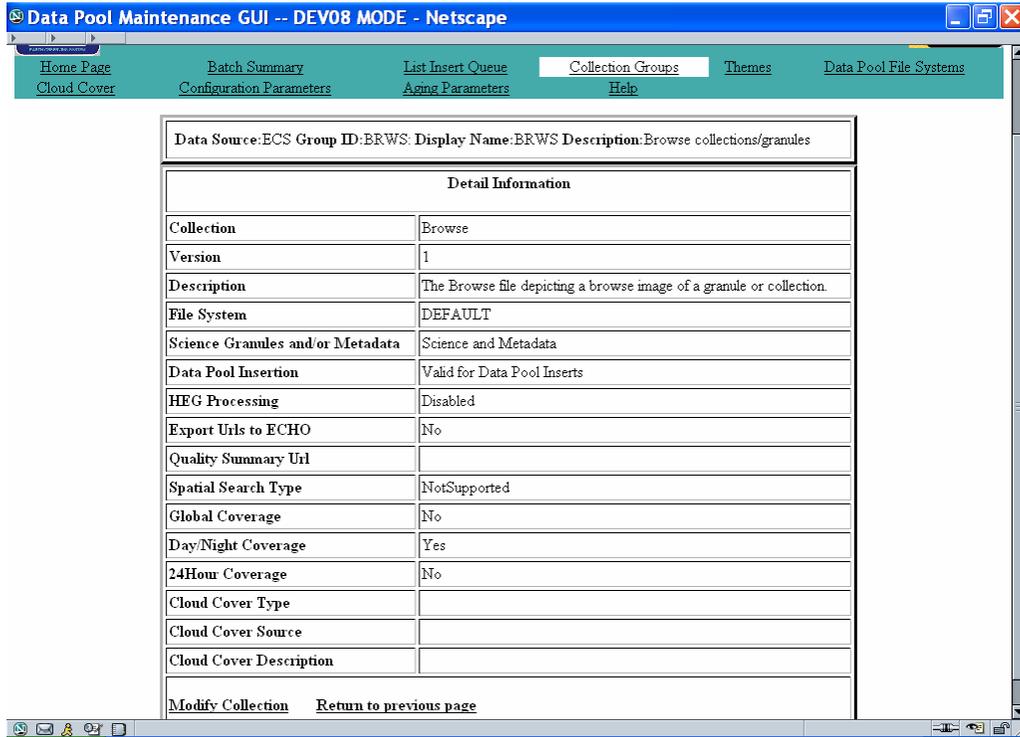


Figure 17.9-16. Collection Detail Information Page

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. Figure 17.9-17 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.

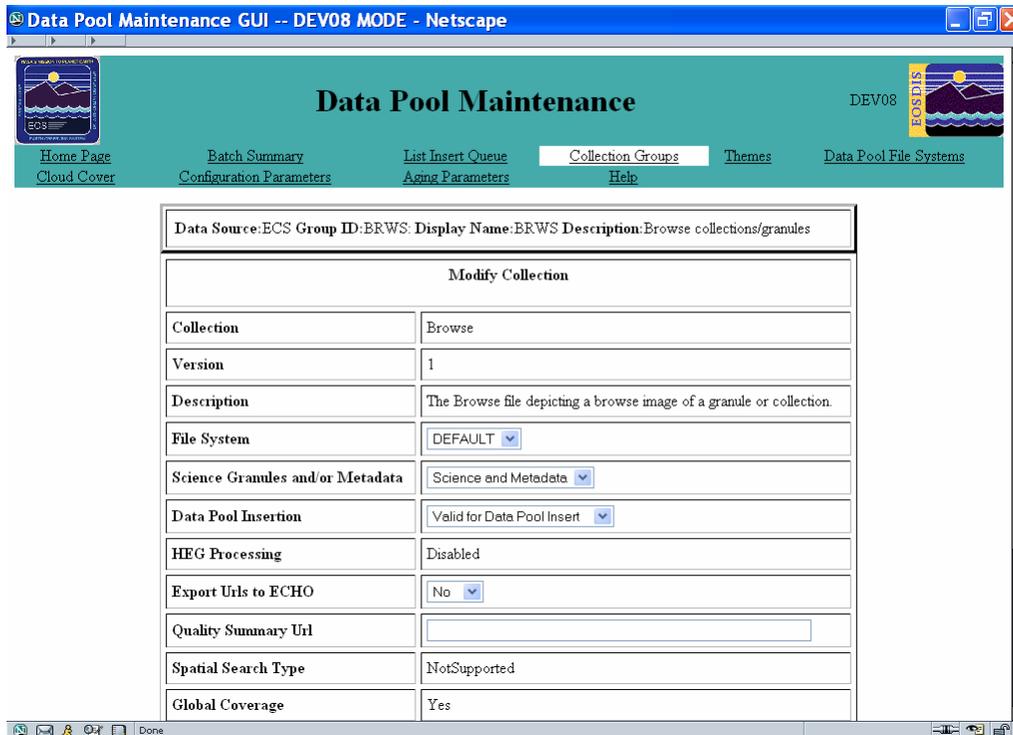
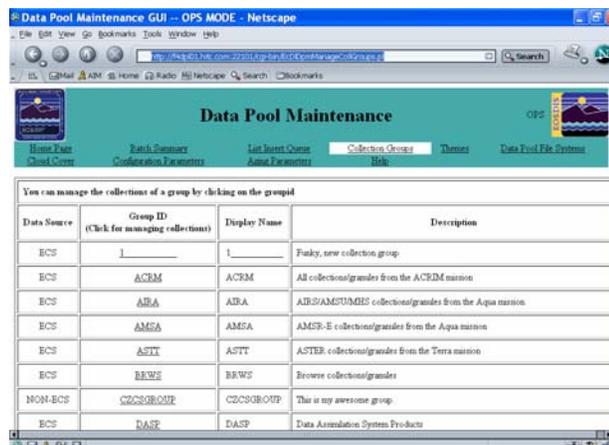


Figure 17.9-17. Modify Collection Page

17.9.8.1 View Collection Groups

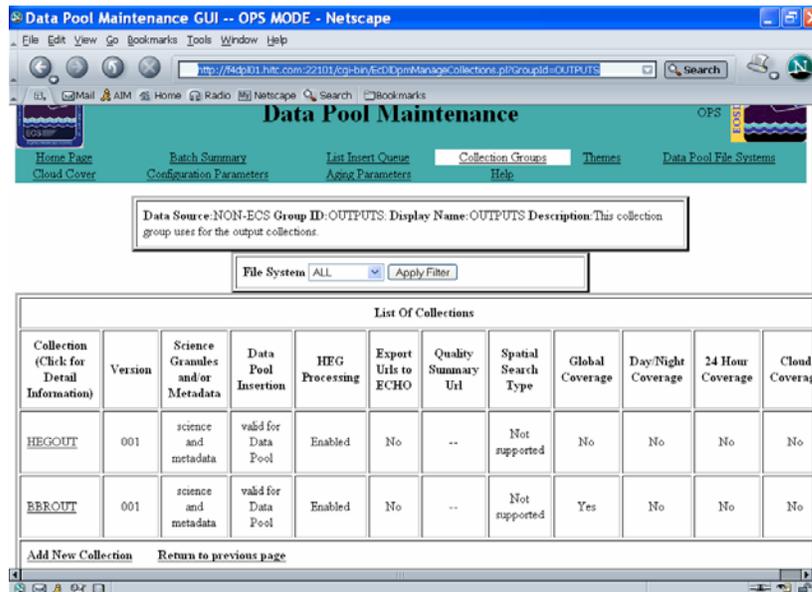
- 1 Launch the DPM GUI.
 - The DPM GUI Home Page is displayed.
- 2 Click on the **Collection Groups** link.
 - The Collection Group page is displayed.



- 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.**
 - 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.

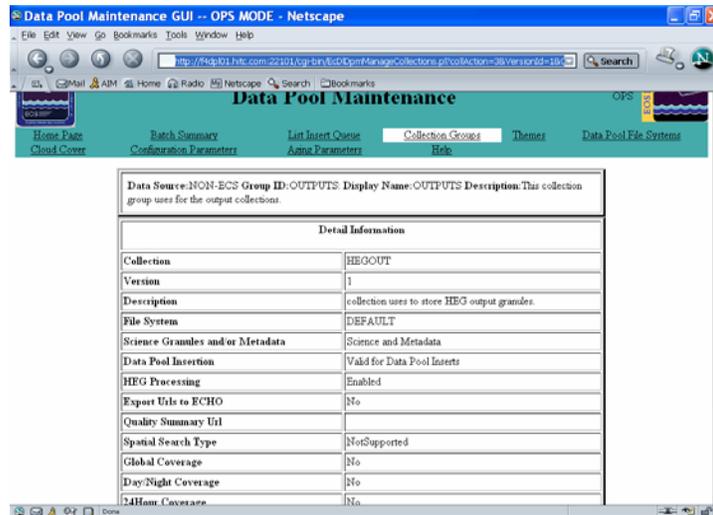


- 5 Observe data displayed on the **List of Collections** page.
- Near the top of the **List of Collections** page contains the following basic collection group information:
 - Data Source.**
 - Group ID.**
 - Display Name.**

- **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.**
 - 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 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.**
 - **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 17.9-17 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.

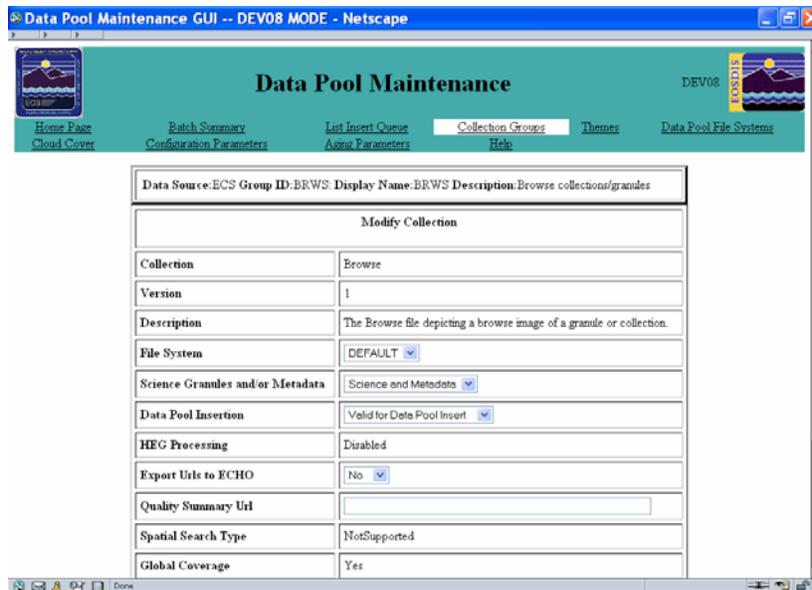


Figure 17.9-17. Modify Collection Page

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 modify/add a collection group:

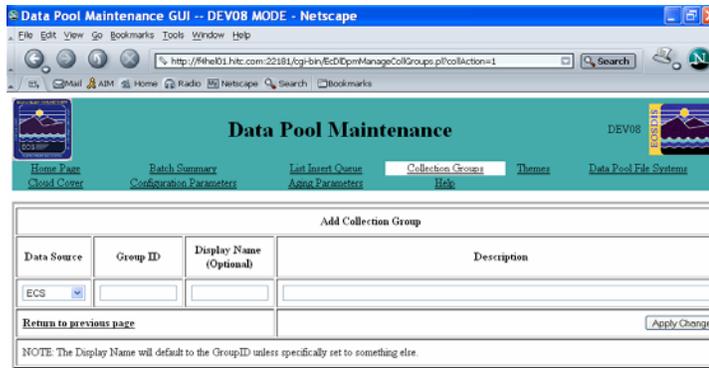
NOTE: Although the following procedure is applicable, most of the time new collection groups will be added only during releases of new software versions and you will not use this procedure often.

Caution

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

17.9.8.3 Add a Collection Group

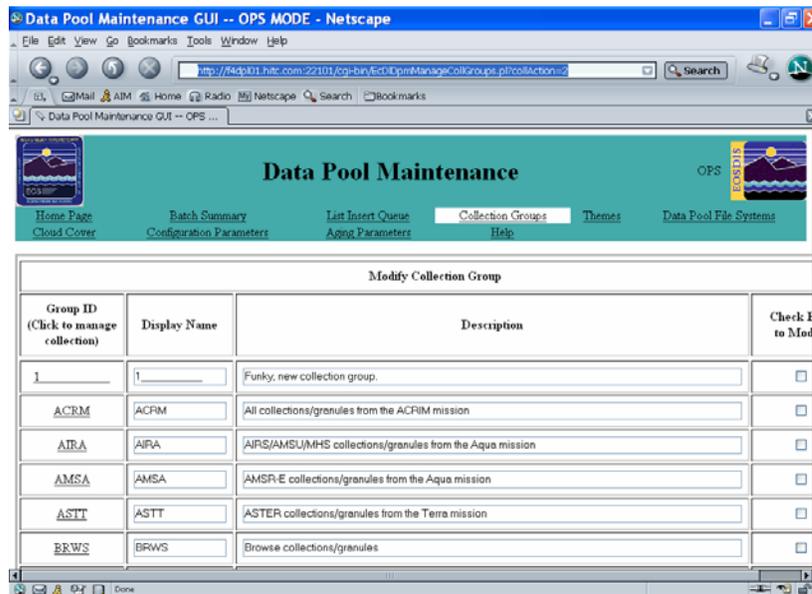
- 1** Launch the **DPM GUI**.
 - The **DPM GUI Home Page** is displayed.
- 2** Click on the **Collection Groups** link.
 - The Collection Groups page is displayed.
- 3** Click on the **Add Collection Group** link at the bottom of the page.
 - The **Add Collection Group** page is displayed providing a page with three columns of text-entry fields, **Data Source**, **Group ID**, **Display Name**, and **Description**.



- 4 Select the **Data Source** from the pull-down window.
 - **ECS** or **NONECS** is displayed in the 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.

17.9.8.2 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 five columns: **Data Source**, **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.

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 17.9-18) 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 17.8-19) 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.

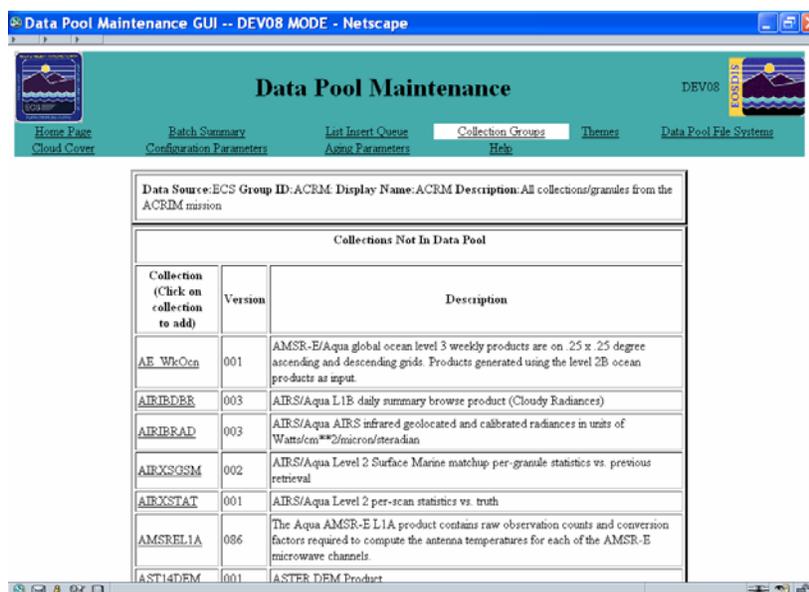


Figure 17.9-18. Collections Not in Data Pool Page

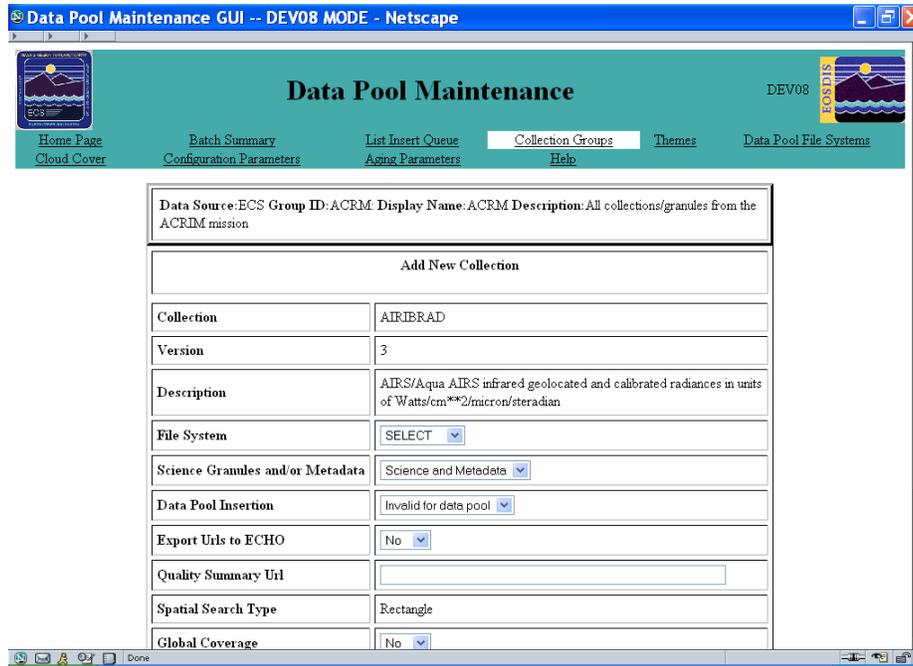
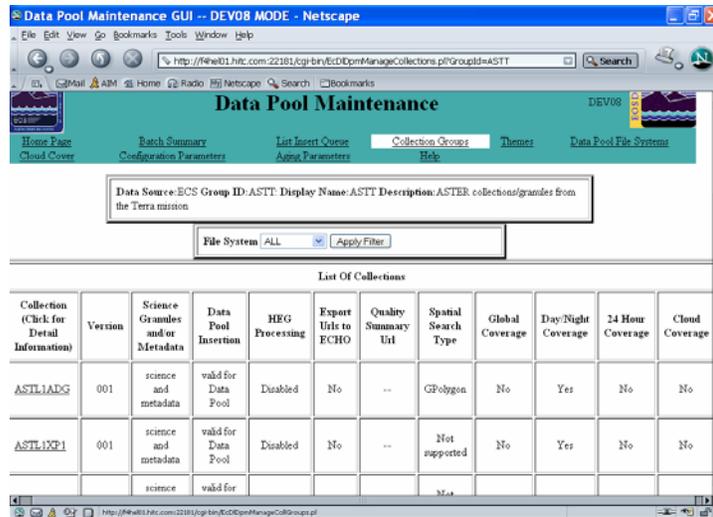


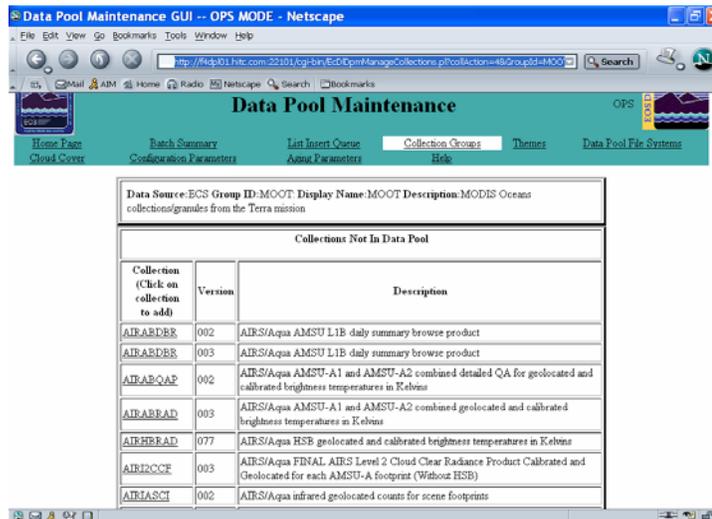
Figure 17.9-19. Add New [ECS] Collection Page

17.9.8.4 Add an ECS Collection to a Collection Group

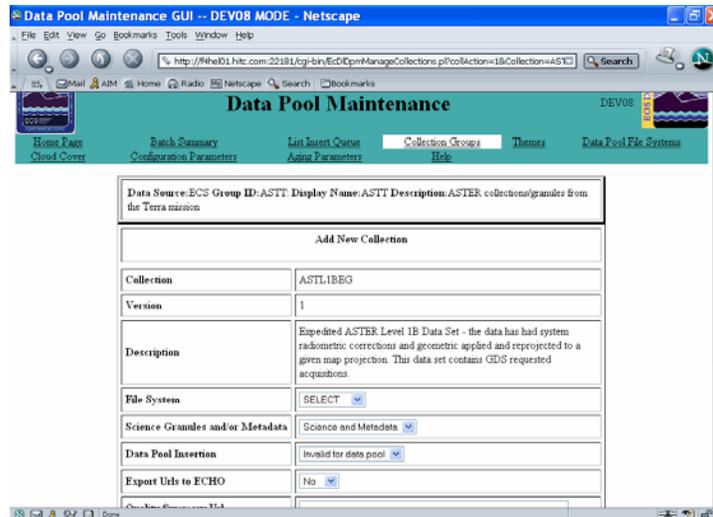
- 1 Launch the **DPM GUI**.
 - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Collection Groups** link.
 - The Collection Group page is displayed.
- 3 Click on the **Group ID** link for the ECS collection group to which the collection is to be added.
 - The **List of Collections** page is 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.



- 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.



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 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.
- 11 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.

- 12 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.
 - 13 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.
 - 14 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).
 - 15 To view details of cloud cover type and source, click on the **View Details** link adjacent to the **Cloud Cover Type & Source** option list.
 - 16 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 17.9-20 shows the **Modify Collection Group** page that a full-capability operator obtains by clicking on the **Modify Collection Group** link at the bottom of the Collection Groups page. On the **Modify Collection Group** page, the full-capability operator can modify the Display Name and/or Description of one or more collection groups then mark the group for change by checking the box(es) in the last column of the table. The operator implements the change(s) with a click on the **Apply Change** button at the bottom of the page.

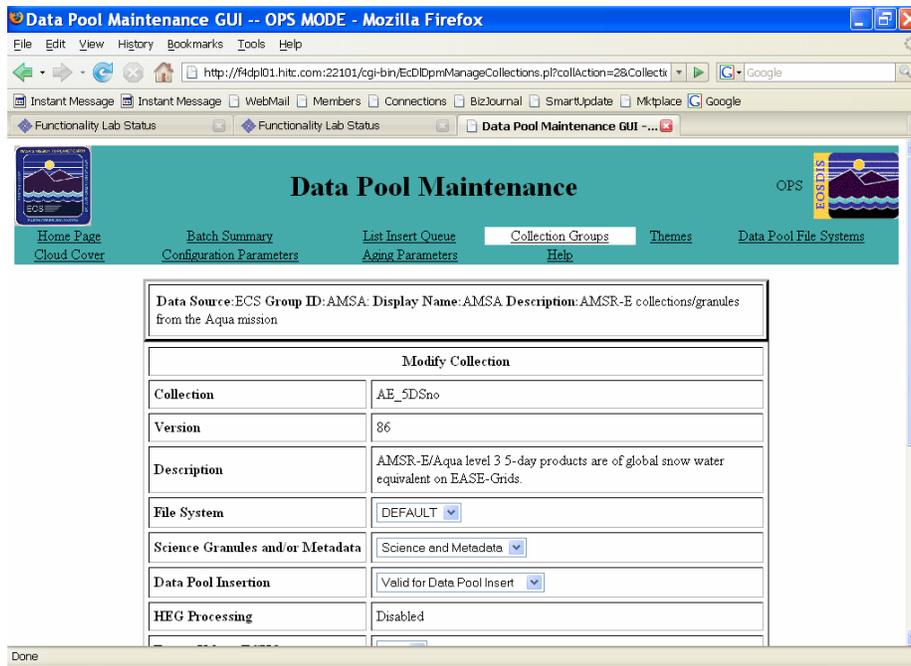
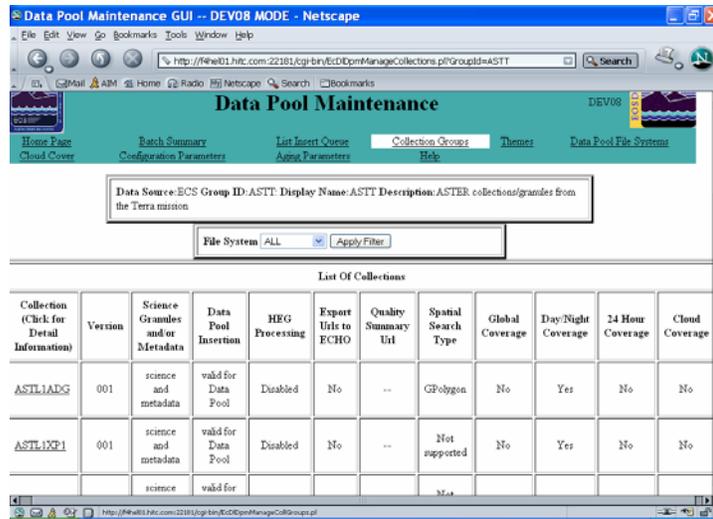


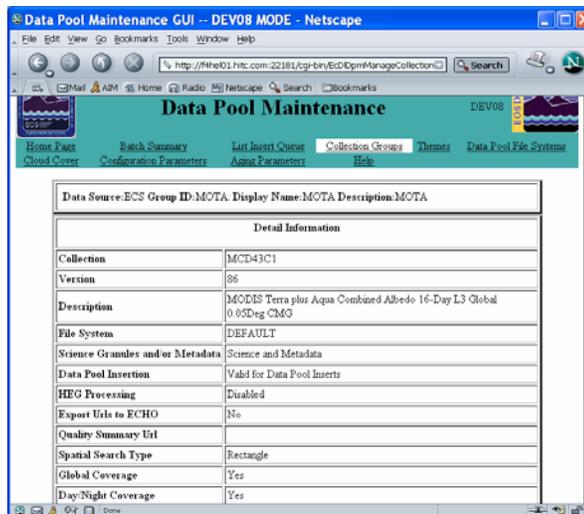
Figure 17.9-20. Modify Collection Page

17.9.8.5 Modify an ECS Collection

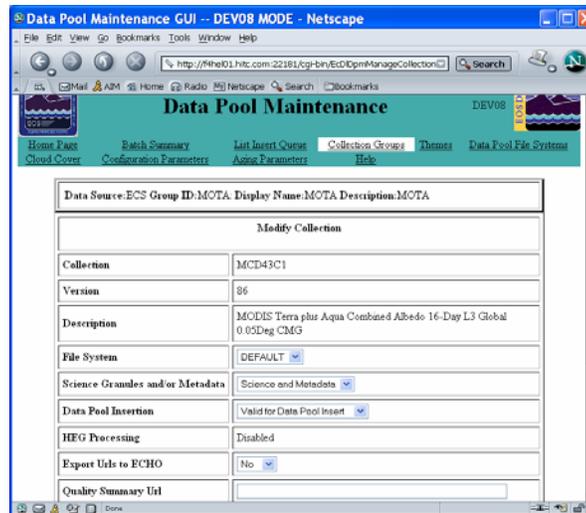
- 1 Launch the **DPM GUI**.
 - The **DPM GUI Home Page** is displayed.
- 2 Click on the **Collection Groups** link.
 - The Collection Group page is displayed.
- 3 Click on the **Group ID** link for the ECS collection group to which the collection is to be added.
 - The **List of Collections** page is displayed.



- 4 Click on the desired link found in the **Collection (Click on collection to add)** column.
- The **Detail Information** page is displayed.



- 5 Click on the **Modify Collection** link.
- The **Modify Collection** page is displayed.



NOTE: On the ECS collection version of the **Modify Collection** page, the **Collection**, **Version**, **Description**, **Spatial Search Type**, and **HEG Processing** 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 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 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.
- 11 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.

- 12 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.
 - 13 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.
 - 14 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).
 - 15 To view details of cloud cover type and source, click on the **View Details** link adjacent to the **Cloud Cover Type & Source** option list.
 - 16 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.
-

17.9.9 Themes

Figure 17.9-21 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 17.9-21).

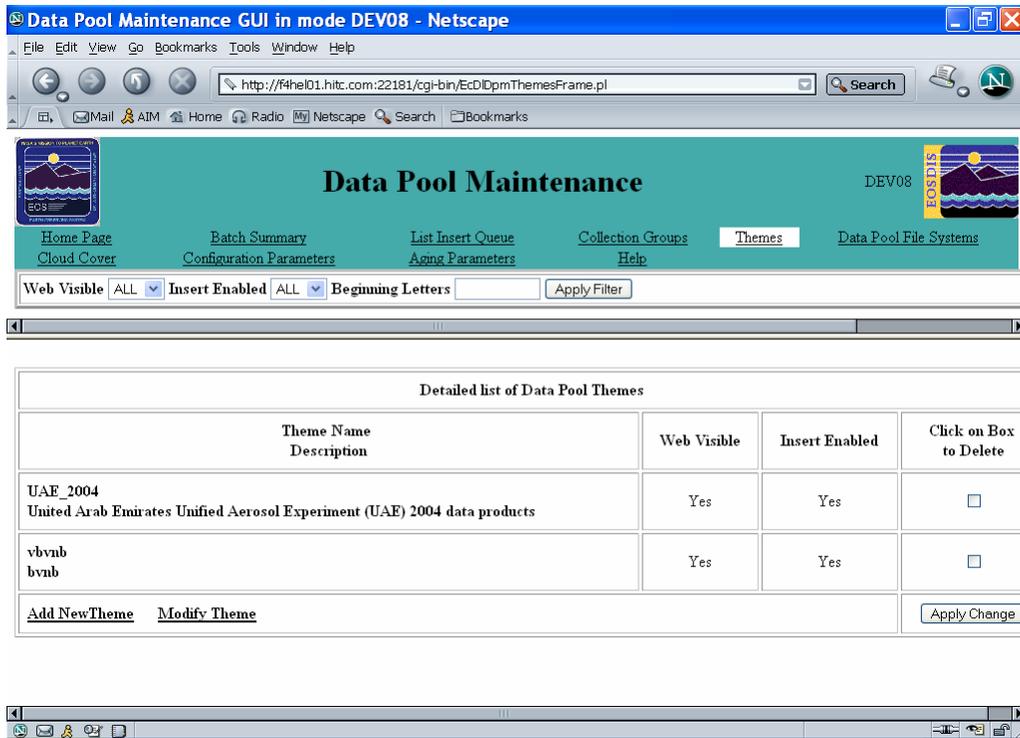


Figure 17.9-21. Detailed List of Data Pool Themes Page

If a full-capability operator clicks on the **Add New Theme** link of the **Detailed List of Data Pool Themes** page shown in Figure 17.9-21, the **Add New Theme** page (Figure 17.9-22) is displayed. To specify a theme, the operator types 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 17.9-21.

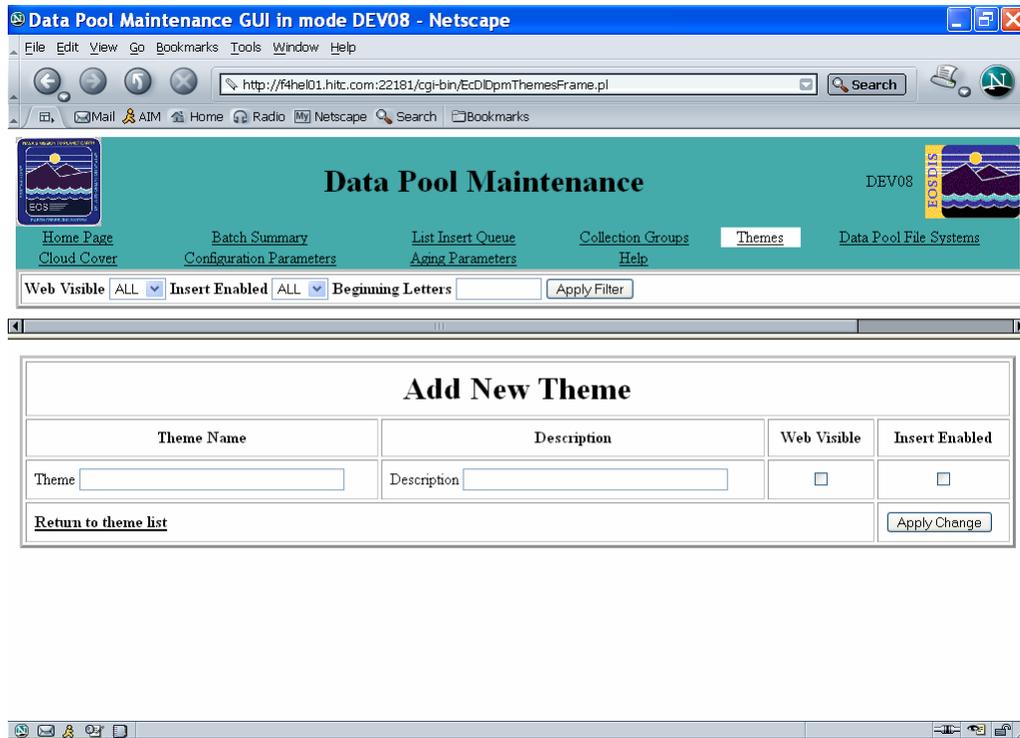
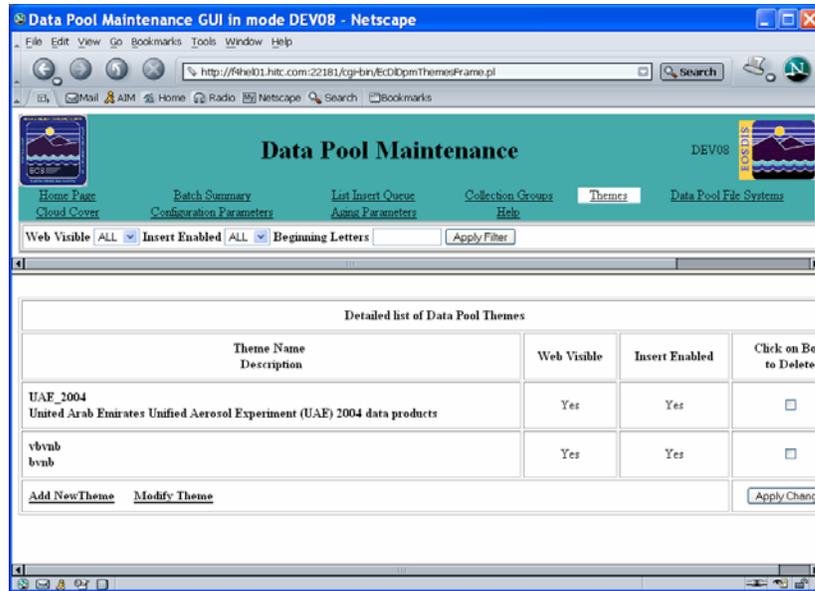


Figure 17.9-22. Add New Theme Page

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.

17.9.9.1 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.



- 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.

17.9.9.2 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 17.9-21, the **Modify Theme** page (Figure 17.9-23) 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 17.9-21.

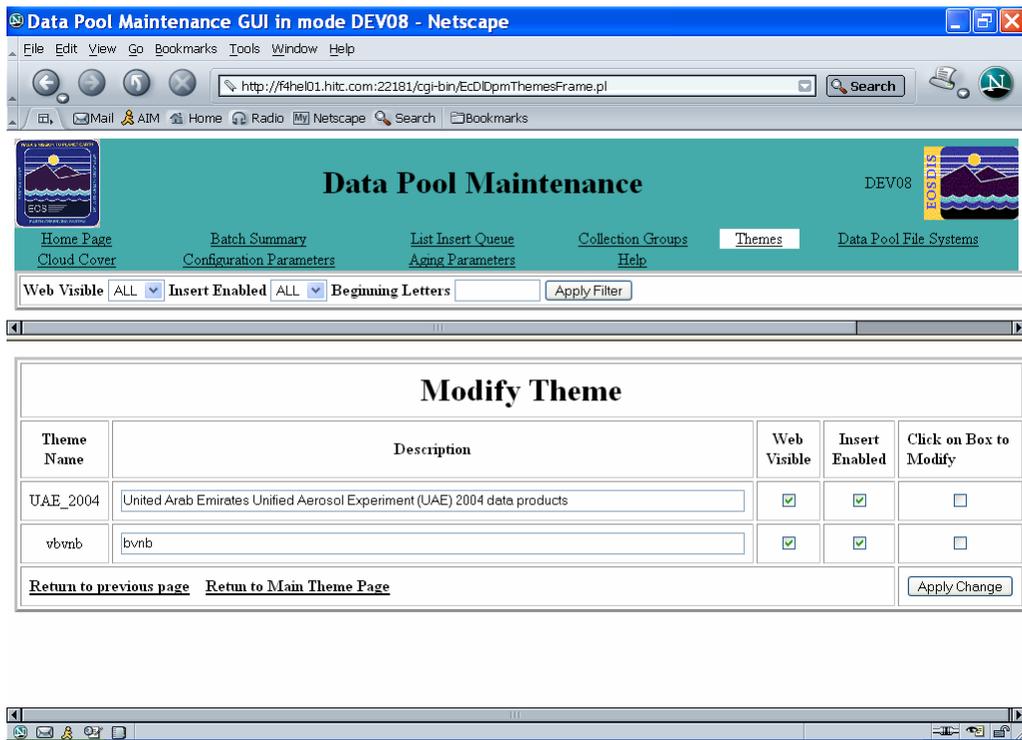
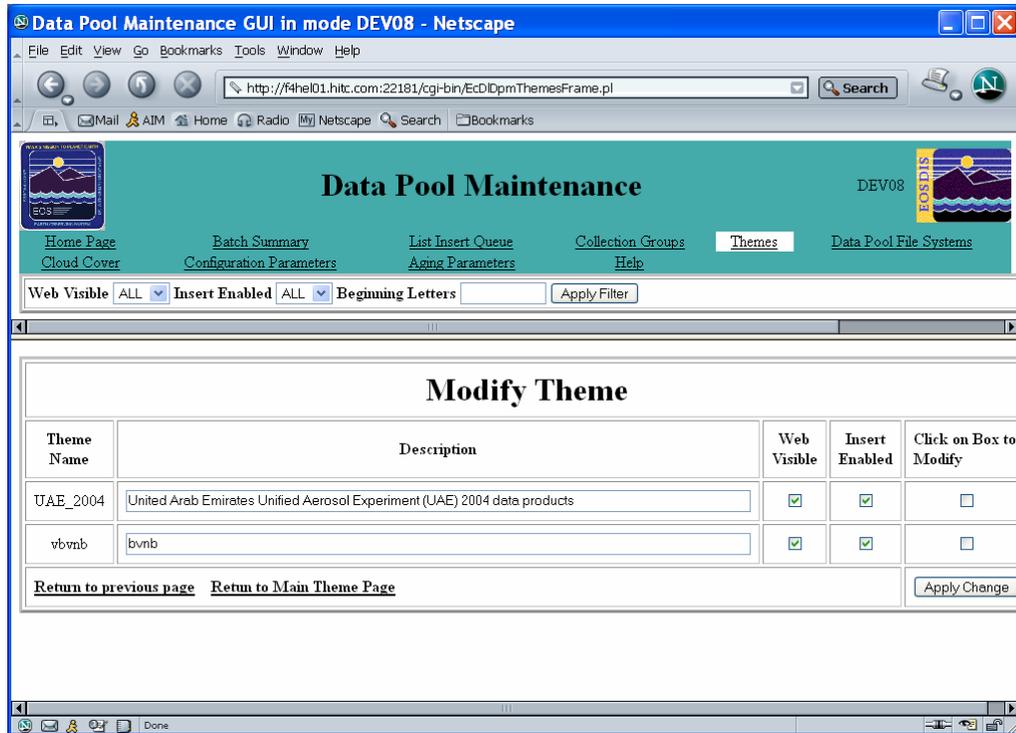


Figure 17.9-23. Modify Theme Page

17.9.9.3 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.

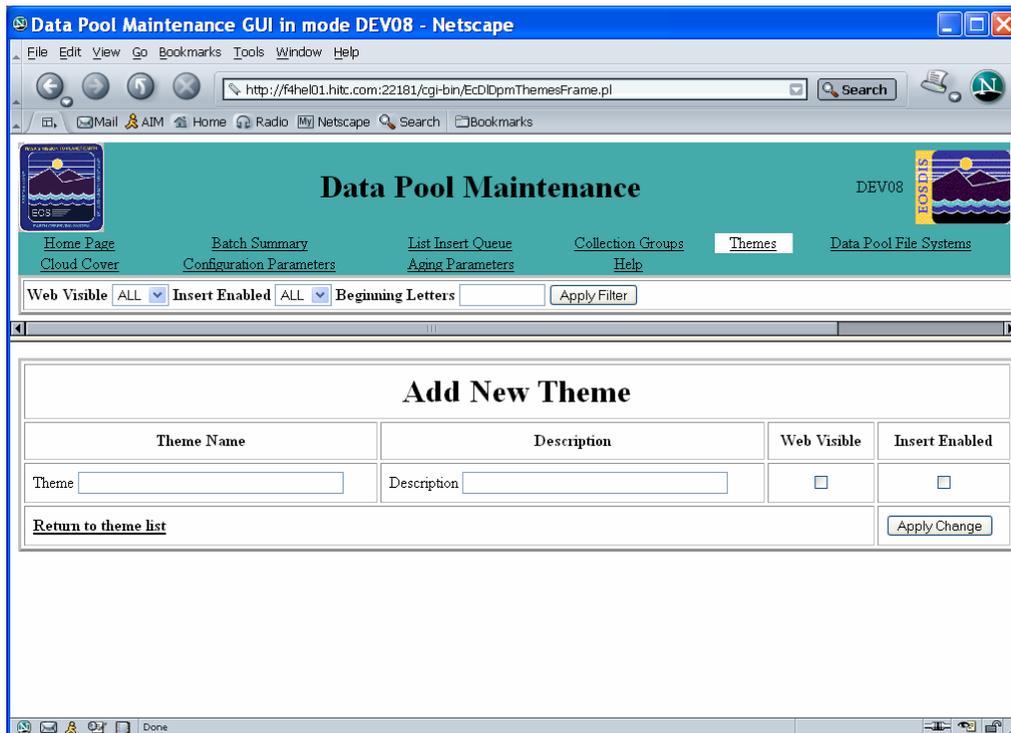


- 4 To filter data displayed on the **Modify Theme** page use the **Filter a List of Themes** procedure.
- 5 To change the description of a theme (if applicable) type the desired description in the **Description** field for the theme name.
 - The **Description** may have no more than 255 characters.
- 6 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.
- 7 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.
- 8 Repeat Steps 5 through 7 as necessary for any additional themes to be modified.
- 9 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:

17.9.9.4 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.



- There are theme filters (and associated **Apply Filter** button) for displaying data on the **Modify Theme** page.
 - The filters serve no real function on this page (there is nothing to filter).
 - 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:

17.9.9.5 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 To filter data displayed on the **Detailed List of Data Pool Themes** page use the **Filter a List of Themes** procedure (previous section of this lesson).
 - 4 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.
 - 5 Repeat Step 4 as necessary for any additional themes to be deleted.
 - 6 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.
-

17.9.10 Help

Figure 17.9-24 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**.

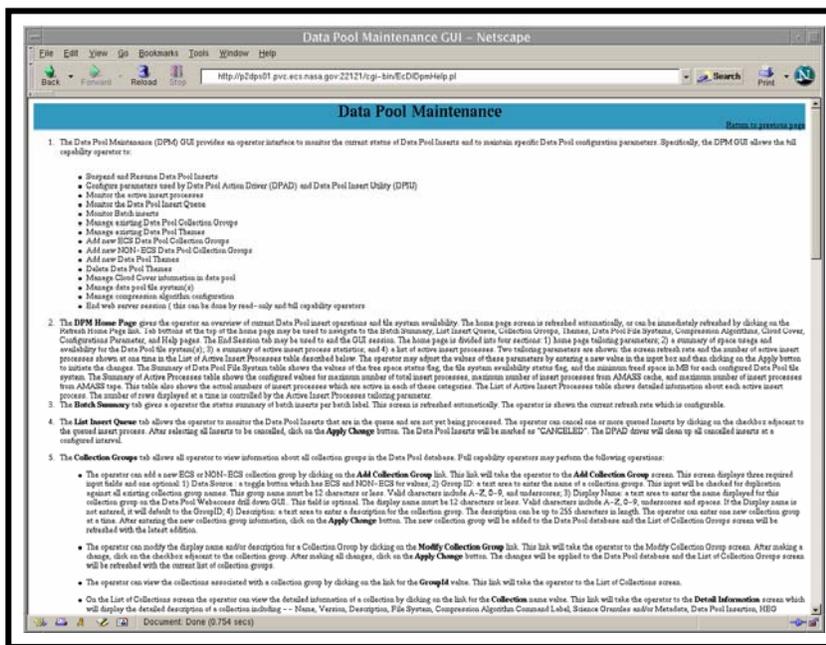


Figure 17.9-24. Help Page

17.10 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.
- **Data Pool Cleanup Utility:** a script to remove expired granules from Data Pool disks and inventory and free up space in the Data Pool. It permits unlinking granules from a thematic collection, and permits data pool removal of granules that belong to a given thematic collection if they meet the cleanup criteria specified by the other command line parameters.

- 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.
- Data Pool Archive Access Statistics Data Utility: an operational support tool for archiving the granule access data from the Data Pool database into a tab-delimited ASCII file.
- Data Pool Delete Access Statistics Data Utility: an operational support tool for deleting granule access data in a specified time range from the Data Pool database.
- Data Pool Restore Access Statistics Data Utility: an operational support tool for restoring granule access data for a specific time range from an ASCII archive file to the Data Pool database.
- Batch Insert Utility: a command-line utility for inserting non-ECS data and ECS data that are already in the archive into the Data Pool.
- 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 QA Update Utility: a command-line interface for updating the QA data for granules in the Data Pool inventory. Both the inventory and the corresponding XML files on disk are updated.
- 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.
- Non-ECS Action Driver: a command line tool used to insert Non-ECS data into the Data Pool.

Table 17.10-1 provides an Activity Checklist for Data Pool Scripts addressed in this section.

Table 17.10-1. Data Pool Scripts - Activity Checklist

Order	Role	Task	Section	Complete?
1	Archive Technician	Extend the Retention for Selected Science Granules Using the Update Granule Utility	(P) 17.10.1.1	
2	Archive Technician	Invoke the Data Pool Cleanup Utility Manually	(P) 17.10.2.1	
3	Archive Technician	Establish Data Pool Cleanup to Run with Cron	(P) 17.10.2.2	
4	Archive Technician	Specify Data Pool Access Statistics Rollup Start Time and DPASU Execution with <i>cron</i>	(P) 17.10.3.1	
5	Archive Technician	Specify Data Pool Access Statistics Utility Execution from the Command Line	(P) 17.10.3.2	
6	Archive Technician	Archive Access Statistics Using the Data Pool Archive Access Statistics Data Utility	(P) 17.10.3.3	
7	Archive Technician	Delete Access Statistics Using the Data Pool Archive Access Statistics Data Utility	(P) 17.10.3.4	
8	Archive Technician	Restore Access Statistics Using the Data Pool Archive Access Statistics Data Utility	(P) 17.10.3.5	
9	Archive Technician	Perform Batch Insert of Data Into the Data Pool	(P) 17.10.4.1	
10	Archive Technician	Running the Bulk Metadata Generation Tool (BMGT)	(P) 17.10.5.1	
11	Archive Technician	Running the Bulk URL Utility	(P) 17.10.6.1	
12	Archive Technician	Running the Most Recent Data Pool Inserts Utility	(P) 17.10.7.1	
13	Archive Technician	Running the Data Pool Collection-to-Group Remapping Utility	(P) 17.10.8.1	
14	Archive Technician	Running the QA UpdateUtility	(P) 17.10.9.1	
15	Archive Technician	Running the Data Pool Move Collections Utility	(P) 17.10.1	
16	Archive Technician	Running the Data Pool Hidden Scrambler Utility in Rename Mode	(P) 17.10.11.1	
17	Archive Technician	Running the Non-ECS Action Driver Insert Tool	(P) 17.10.12.1	

17.10.1 Extending the Period of Retention for Granules in the Data Pool

We have seen that a change in user interest in data from a particular location may arise because of unusual circumstances (e.g., weather, natural event) and that as a result it may be desirable to

extend the period of retention in a Data Pool insert subscription. Such circumstances may also make it desirable to retain certain data already in the Data Pool for a longer period of time than originally specified. Data Pool maintenance personnel can run the Update Granule Utility to update the expiration date for selected 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 granule may be updated using manual input. Multiple 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.

17.10.1.1 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 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 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 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 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 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 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 Data Pool 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 **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.
-

17.10.2 Running the Data Pool Cleanup Utility

The Data Pool Cleanup Utility permits DAAC Operations Staff to remove expired granules from the Data Pool disks and inventory. In addition, the Data Pool Cleanup Utility reports (via an external utility) to the EOS Clearing House (ECHO) the granules that are to be (or that have been) deleted.

The Cleanup Utility may be executed using a **-noprompt** argument to suppress all confirmations and warnings normally displayed to standard output.

The Cleanup Utility must be executed on the machine where the granules are located. Qualification for cleanup is based on three criteria: expiration date/time, retention priority, and theme association. Figure 17.10-1 illustrates how the utility uses expiration date/time and retention priority in combination to select granules for removal from the Data Pool. The operator specifies values for these criteria and, if desired, themes and cross-references, to control cleanup operations. It is also possible to use the utility to validate the Data Pool inventory and disk cache by checking for and removing *orphans* (disk files not associated with any entry in the Data Pool inventory) and *phantoms* (entries in the Data Pool inventory that refer to files that no longer exist on the Data Pool disks).

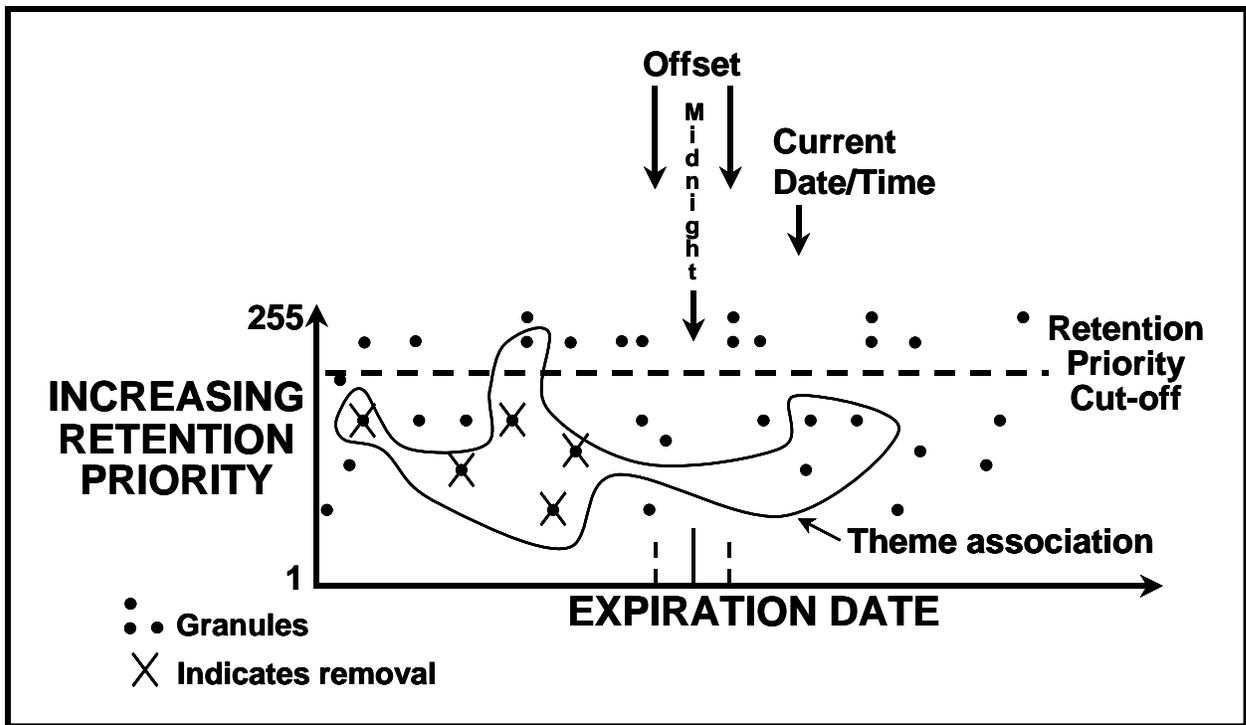


Figure 17.10-1. The Data Pool Cleanup Utility Uses Expiration Date, Retention Priority, and/or Theme Association to Qualify Granules, Marked with X, for Deletion

To determine whether a granule qualifies for deletion, the utility first compares the granule's expiration date (insert date plus retention period in days specified in the insert subscription) with a cut-off date/time. If a granule's expiration date is prior to the cut-off, the granule qualifies as expired.

- The default cut-off date/time is set to midnight of the previous day.
- The operator is permitted to specify an offset (from the previous midnight) in hours to add or subtract hours to determine a cut-off date/time for deletion.
 - For example, **-offset -5** would delete all granules that had expired as of 7:00 P.M. yesterday.

Next, the utility compares the granule's retention priority with any priority **limit** the operator has specified to identify those granules that should be retained in the Data Pool even though their expiration date has passed.

- Retention priority is an integer from 1 to 255.

- For example, **-limit 150** would delete all granules with priority less than or equal to 150.
- Retention priority for granules already in the Data Pool may be modified using the granule expiration update script.

The Data Pool Cleanup Utility removes those granules with expiration date prior to the cut-off date/time and with retention priority less than or equal to the specified limit. If a priority limit is not specified in command-line input parameters at the time it is invoked, the Cleanup Utility reads the parameter 'DEFAULT_LIMIT' from its configuration file to get a priority limit. If the operator does not wish to use retention priority as a criterion for deletion, the default limit should be set to 255. If the operator specifies a theme name, the utility applies the removal criteria only to those granules associated with the theme.

The Cleanup Utility can alternatively take as input a file listing the granuleId for each granule to be deleted. The file can contain one granuleId or more than one granuleId per line separated by white space. The **-file** option may not be used with any other options other than the **-noprompt** option.

Another file option is **-geoidfile** (e.g., **-geoidfile geoid20040304**). It specifies the name of a file containing GeoIDs, which are a combination of science type, ESDT short name and version ID, and ECS Science Data Server database ID. Granules in the file whose ECS ID match those in the Data Pool are candidates for Data Pool cleanup if specified by this option. The **-geoidfile** option may not be used in conjunction with any other options other than the **-noprompt** option. Note that the GeoID file can contain science granules as well as non-science granules because the science data server may delete these types of granules. The input value for this parameter is logically defined to be the output of any Science Data Server Phase 1 (EcDsBulkDelete.pl) granule deletion run. This causes the Data Pool cleanup utility to clean up any Science Data Server granules found in the GeoID input file to be removed from the Data Pool database.

The Cleanup Utility cleans up non-ECS data just as it does ECS data. It can remove granule cross references associated with a given theme, and also remove the granules associated with the theme. The option **-themexref** specifies a theme for which all cross-references are to be removed from the Data Pool. The option **-theme** specifies a theme for which associated granules are to be removed. If a granule is referenced to more than one theme, the **-theme** option removes only the cross-reference to the specified theme, without removing the granule. The theme name must be enclosed in quotes (e.g., **-theme "Ocean Temperatures"** or **-themexref "Surface Reflectance"**). The **-themexref** option cannot be used with any options other than the **-noprompt** option.

The **-ecsgrandel** option indicates that only granules removed in the ECS system from the Science Data Server inventory are to be removed from the Data Pool if they exist. The option may not be used in conjunction with any options other than the **-noprompt** option. No other cleanup occurs when the **-ecsgrandel** option is specified.

The **-echomode** parameter specifies the method by which the Cleanup Utility reports deletion candidates to ECHO. The **-echomode** parameter takes one of three values; i.e., **predelete**, **finishdelete** or **deleteall**.

When **predelete** is specified, the Cleanup Utility builds the list of items to clean up from the Data Pool and reports them to ECHO through the EcBmBulkURL utility. No data is actually cleaned up from the disks or database inventory using **predelete**.

When **finishdelete** is specified as the value for **-echomode**, the Cleanup Utility deletes all of the data that was last found during a run with the **predelete** parameter. The difference is that the EcBmBulkURL utility is not invoked because this run performs the actual cleanup of the database inventory and disks of what was presumably reported to ECHO during the previous run.

When **deleteall** is specified as the value for **-echomode**, the Cleanup Utility builds its list of items to clean up, actually cleans them up, and notifies ECHO via the EcBmBulkURL utility. The **deleteall** value does not allow for a time lag between the Cleanup Utility deleting the granules and ECHO performing its own clean up of URLs.

The normal sequence for cleanup is to run the Cleanup Utility twice: specifying **predelete** for the first run and **finishdelete** for the second run. Note that an **-echomode** parameter with a value of **finishdelete** can only be specified by itself because the list of items to delete will have already been determined by the previous (**predelete**) run.

If a **predelete** run is performed, the subsequent run *must* specify **finishdelete** in order to perform the actual deletions. The Cleanup Utility enforces the requirement to avoid operator error. The **predelete/finishdelete** run sequence can be viewed as a logical run done in two parts. The values of **predelete** and **deleteall** may be used with each of the other parameters specific to performing Data Pool Cleanup except **themexref**.

There are three types of runs that can be performed with the Cleanup Utility:

- Cleanup only.
- Validation only.
- Cleanup followed by validation.

When involved in “cleanup” processing, the Cleanup Utility performs the following actions:

- Removes from the Data Pool disks and inventory all Data Pool granules, associated browse files, and browse links that meet the specified cleanup criteria (provided that no other granules are cross-referenced to them – i.e., linked by a theme). This occurs when the **-echomode** parameter has a value of **finishdelete** or **deleteall**. (No actual deletion occurs during **predelete**.)
- Removes all recent insert files (with names prefixed with **DPREcentInsert**) that are older than seven days. The relevant files are found in `/datapool/<mode>/user/<fs1>` and `/datapool/<mode>/user/<fs1>/<group>/<esdt>`.
- Exports a list of deleted granules for ECHO accessibility by invoking an external utility (i.e., **EcBmBulkURLStart**) when the Cleanup Utility **-echomode** parameter has a value of either **finishdelete** or **deleteall**.
 - If there are granules being deleted that qualify for ECHO export, the Cleanup Utility generates an XML file containing a list of those granules and stores it in

the /datapool/<mode>/user/URLExport directory for files that are ftp pulled and ftp pushes files when Bulk URL is configured to ftp push the data to ECHO.

- If the Data Pool Cleanup Utility is run in -echomode finishdelete, the EcBmBulkURLStart utility is not called.
- Removes all HEG conversion files associated with the HEG order IDs that have the status of "DONE" or "FAILED" and a timestamp older than a certain cleanup age.
 - The HEG order IDs are provided in the DIOrder table and the cleanup age is specified by the "HEGCleanupAge" parameter in the DIConfig table of the Data Pool database.
 - The HEG conversion files for each order ID are stored in the /datapool/<mode>/user/downloads/<orderID> directory. (HEG orders and conversion files are generated when end users request HEG-converted data using the Data Pool Web Access tool.)

When involved in “validation” processing, the Data Pool Cleanup Utility performs the following actions:

- Validates the Data Pool inventory and disk content by checking for the existence of orphans and/or phantoms and either removing them or just logging them depending on the command line options specified.

Validation requires either the **-orphan** parameter or the **-phantom** parameter or both. The **-orphan** parameter finds/removes data in the Data Pool that is not represented by entries in the Data Pool inventory. The **-phantom** parameter finds/removes entries in the Data Pool inventory that have one or more science or metadata files, or associated browse files, missing from the Data Pool. To specify just logging of the discrepancies, the operator uses the option **-nofix**.

The **-maxorphanage** validation option specifies the maximum orphan age in days (e.g., **-maxorphanage 5**). The value specified must be greater than or equal to three days. The Data Pool inventory validation function will consider only those files on disk as orphans whose age is equal to or larger than the maximum orphan age specified. If the parameter is omitted, the default value specified in the configuration file is used.

The **-collgroup** validation option limits the Data Pool validation to the specified collection group(s). Single or multiple collection groups can be specified on the command line. If multiple collection groups are specified, they must be separated by commas, with the string enclosed in double quotes (e.g., “MOAT, ASTT”). By default all collection groups in the Data Pool inventory are included in the validation if the **-collgroup** option is not specified.

A validation run can be time-consuming and should not be run as often as the cleanup runs, because it potentially involves the checking of all files in the entire Data Pool inventory against those on the Data Pool disk in order to find and remove the discrepancies. It is advised that the validation function be run using the **-collgroup** option whenever possible to limit the validation to a limited number of collection groups.

If the Cleanup Utility is interrupted during execution, upon restart it continues from the point of interruption. Furthermore, in the interest of low database contention, the Cleanup Utility allows only one instance of itself to execute concurrently.

17.10.2.1 Invoke the Data Pool Cleanup Utility Manually

- 1 Log in at the machine where the Data Pool Cleanup 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 Cleanup 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**.

NOTE: There are three types of runs that can be performed with the Cleanup Utility; i.e., “cleanup only,” “validation only,” or “cleanup followed by validation.”

NOTE: The normal sequence for cleanup is to run the Cleanup Utility twice: specifying **predelete** for the first run and **finishdelete** for the second run. Note that an **-echomode** parameter with a value of **finishdelete** can only be specified by itself because the list of items to delete will have already been determined by the previous (**predelete**) run.

NOTE: If a **predelete** run is performed, the subsequent run *must* specify **finishdelete** in order to perform the actual deletions. The Cleanup Utility enforces that requirement to avoid operator error. The **predelete/finishdelete** run sequence can be viewed as a logical run done in two parts.

- 3 To perform a “cleanup only” run, at the UNIX prompt enter:
EcDlCleanupDataPool.pl <MODE> -echomode <echomode> [-noprompt] [-offset <hours>] [-limit <priority>] [-theme <themeName>] [-nodmap]
OR
EcDlCleanupDataPool.pl <MODE> -echomode <echomode> [-noprompt] -file <fileName>
OR
EcDlCleanupDataPool.pl <MODE> -echomode <echomode> [-noprompt] -geoidfile <fileName>
OR
EcDlCleanupDataPool.pl <MODE> -echomode <echomode> [-noprompt] -ecsgrandel
OR
EcDlCleanupDataPool.pl <MODE> [-noprompt] -themexref <themeName>
 - **<echomode>** is the value specified for the ECHO mode. The value is either **predelete**, **finishdelete**, or **deleteall** (e.g., **-echomode predelete**).
 - **<hours>** is the value of the offset. It can be either a positive number (e.g., **-offset 2**) or a negative number (e.g., **-offset -5**). If the **-offset** option is not specified, the Cleanup Utility uses the default value of 0 (zero).

- <priority> is the value of the priority limit. It is a number from 1 through 255 (e.g., -limit 200). If the -limit option is not specified, the Cleanup Utility uses the default value specified in the configuration file.
- <themeName> is the name of a theme to be associated with either the -theme option or the -themexref option. The name of the theme must be in quotes (e.g., -theme "Ocean Temperatures" or -themexref "Surface Reflectance").
- <fileName> is the name of a file to be associated with either the -file option or the -geoidfile option (e.g., -file clean20040404 or -geoidfile geoid20040304). The file will be read by the Cleanup Utility to determine what granules to clean up.
- The -ecsgrandel option indicates that only granules removed in the ECS system from the Science Data Server inventory will be removed from the Data Pool if they exist. No other cleanup occurs.
- The -nodmap option is used to avoid updating the current density map for each collection having granuleIds in DIFilesToBeDeleted.
- The Cleanup Utility runs and the Cleanup Utility log file EcDICleanup.log records errors, warnings, and information about utility events.

4 To perform a "validation only" run, at the UNIX prompt enter:

- EcDICleanupDataPool.pl <MODE> -orphan | -phantom [-noprompt] [-collgroup <groupList>] [-maxorphanage <days>] [-nofix]**
- For validation either the -orphan parameter or the -phantom parameter or both must be specified.
 - <groupList> is the name of the collection group(s) to be validated (e.g., "MOAT, ASTT"). The collection group(s) must be enclosed in quotes and if there are multiple groups, they must be separated by commas. If the -collgroup option is not specified, all collection groups in the Data Pool inventory are included in the validation.
 - <days> is the number of days (at least 3) after which files on the Data Pool disks are considered orphans if they do not have corresponding entries in the Data Pool inventory. The default value in the configuration file (e.g., 3) is used if the -maxorphanage option is not specified.
 - The -nofix option prevents reconciling any discrepancies found during validation. The validation results are logged.

5 To perform a "cleanup followed by validation" run, at the UNIX prompt enter a command line with valid options from Steps 3 and 4 plus the -cleanvalidate parameter.

- For example:
EcDICleanupDataPool.pl OPS -echomode predelete -offset 5 -limit 200 -orphan -phantom -cleanvalidate

- 6 If **predelete** was specified as the value for the **-echomode** parameter in Step 3 or Step 5, after the Cleanup Utility has run to completion repeat Step 3 (or Step 5) to perform a cleanup using **finishdelete** as the value for the **-echomode** parameter.
-

In normal operations, the Cleanup Utility is run once a day as a cron job as a "cleanup only" run executing in echo mode of **predelete**. This builds the list of cleanup candidates (based on the expiration date and retention priority) that are reported to ECHO as those that will be deleted in the next run of cleanup. Also, the granules that have been entered in the Science Data Server's deleted granules table will be reported. On a subsequent run within the same 24-hour period, the cleanup utility is run in **finishdelete** mode to perform the actual cleanup processing that was reported to ECHO in the **predelete** mode.

17.10.2.2 Establish Data Pool Cleanup to Run with *cron*

- 1 Log in at a system platform using an account with privileges to remove science, metadata, and browse files from Data Pool disks.
- 2 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.
- 3 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.
- 4 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.
- 5 Type **i** to put the **vi** editor into the insert mode.
 - The **vi** editor is in the insert mode, but no feedback is provided.
- 6 Type the crontab entry, including the appropriate Cleanup Utility command (as described in Section 17.10.32, **Invoke the Data Pool Cleanup Utility Manually**).
 - For example:
**0 1 * * * /usr/ecs/OPS/CUSTOM/utilities/EcDIDataPoolCleanup.pl OPS
-echomode predelete -noprompt**
 - The example would start a **predelete** cleanup run at 1:00 A.M. every day.
- 7 Repeat Step 6 as necessary to enter additional crontab entries, including the appropriate Cleanup Utility command (e.g., to run a **finishdelete** cleanup run at 4:00 A.M. every day).
- 8 Press the **Esc** key.
 - The cursor moves one character to the left and the **vi** editor is in the command mode.

- 9 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.
-

17.10.3 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 **EcDIRollupFtpLogs.pl** runs on a server with access to SYSLOG with FTP access entries; its configuration file is **EcDIRollupFtpLogs.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.

17.10.3.1 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 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 EcDIRollupWebLogs.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
and the cursor is displayed on the first character at the upper left corner of the file.
 - 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
 - If more characters in the time value are to be changed, you can type **x** repeatedly to delete additional characters. For this exercise, you need only delete one character.
- 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**.
- The typed entry appears to the left of the cursor.
- 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., **x4dpl01**).
- 18 To change to the directory containing the configuration file, type the 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 **vi EcDIRollupFtpLogs.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
and the cursor is displayed on the first character at the upper left corner of the file.
 - 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 **EcDIRollupFtpLogs.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/EcDIRollupFtpLogs.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 **EcDIRollupFtpLogs.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 rollover period, using the format **MM/DD**, and may be used to process a previously uncovered period.

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.

17.10.3.3 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., x4oml01).
- 2 To change directory to the directory containing the Data Pool Archive 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 `DIDbArchiveAccessStat <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., x4oml01_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.

17.10.3.4 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., x4oml01).

- 2 To change directory to the directory containing the Data Pool Delete 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 **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., `x4oml01_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.

17.10.3.5 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., `x4oml01`).
- 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 yyyyymmdd, 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., x4oml01_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 DIDbRestoreAccessStat.log records errors, warnings, and information about utility events. The log is written to the directory /usr/ecs/<MODE>/CUSTOM/logs.

17.10.4 Using the Batch Insert Utility

The Batch Insert Utility allows operators to specify Data Pool insert for granules residing in the ECS archive, as well as data from outside ECS (non-ECS granules). The utility queues the granules up for dispatch by the Data Pool Action Dispatcher (DPAD) for insertion by the Data Pool Insert Utility (DPIU). It accepts either a list of ECS granule identifiers or a list of non-ECS names; the list can be provided either as an input file or as standard input. A label identifying a batch of granules is specified as a command-line parameter, using the option **-label**, so that operators can monitor a batch with the DPM GUI.

Granules to be inserted can also be linked to a theme, using the option **-theme**. In fact, the Batch Insert Utility can also be used with that option to link granules already present in the Data Pool to a theme, or to additional themes. However, it is important to note that if the granules were originally inserted into the Data Pool using the Batch Insert Utility, you must use a different batch label when linking the granules to the theme than was used for the original insert. This is necessary because the Batch Insert Utility is designed to reject inserts that are in a batch with a label identical to one for which granules are already being processed. So, even if the batch has been inserted, if the inserts are still in the queue (e.g., with a status of **Completed**), you cannot run another batch with the same label to link them to a theme.

The following procedure is applicable.

17.10.4.1 Perform Batch Insert of Data into the Data Pool

- 1** Log in at the machine where the Data Pool Batch Insert Utility is installed (e.g., x4dpl01).

NOTE: The login must be as either **cmshared** or **allmode** to ensure correct permissions.

- 2 To change to the directory for starting the Batch Insert 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 Batch Insert Utility, in the form **EcDIBatchInsert.pl <MODE> -ecs | -nonecs [-file <pathname>] [options]** then press the **Return/Enter** key.
- The following examples show valid command-line entries for initiating the Batch Insert Utility:
 - **EcDIBatchInsert.pl <MODE> -ecs -file /home/cmshared/<filename>** (to add actions to the action insert queue for all ECS granules specified by granule IDs in the specified file. Because the command does not specify a **-label** parameter, the label is formed from the first 16 characters of the input file name).
 - **EcDIBatchInsert.pl <MODE> -nonecs -file /home/cmshared/<filename> -label Chig_volcano -theme “Chiginagak Volcano 2002”** (to add actions to the insert action queue for all non-ECS granules specified by XML pathnames in the specified input file, with all granules linked with the theme name “**Chiginagak Volcano 2002**” in the Data Pool database).
 - The theme name must already be in the Data Pool database in the DIThemes table; if necessary, use the DPM GUI **Manage Themes** tab to define the theme before running the batch insert.
 - You can use Batch Insert with the **-theme** option to link granules already in the Data Pool to a theme, but if the granules were originally inserted using the Batch Insert Utility, you must use a different batch label than was used for the original insert; otherwise, the insert of the theme links may be rejected.
 - **EcDIBatchInsert.pl <MODE> -ecs -file /home/cmshared/<filename> -mdonly** (to add actions to the action insert queue for all ECS granules specified by granule IDs in the specified file, but insert metadata only. Because the command does not specify a **-label** parameter, the label is formed from the first 16 characters of the input file name).
 - **EcDIBatchInsert.pl <MODE> -ecs -file /home/cmshared/<filename> -rppriority 255** (to add actions to the action insert queue for all ECS granules specified by granule IDs in the specified file, and to set their retention priority to 255. Because the command does not specify a **-label** parameter, the label is formed from the first 16 characters of the input file name).
 - **EcDIBatchInsert.pl <MODE> -ecs -file /home/cmshared/<filename> -rppriority 255 -rperiod 10 -dpriority 5** to add actions to the action insert queue for all ECS granules specified by granule IDs in the specified file, and to set their retention priority to 255 and their retention period to 10 days, with dispatch priority set to 5. Because the command does not specify a **-label** parameter, the label is formed from the first 16 characters of the input file name).

- The Batch Insert Utility runs and events and errors are recorded in the Batch Insert Utility log file EcDIBatchInsert.log.
-

17.10.5 Running the Bulk Metadata Generation Tool (BMGT)

In order to support the development of value-added providers (e.g., IIMS, ESIPs, RESACs, and InfoMarts), ECS sites make an external representation of their metadata holdings available and provide a capability for bulk distribution of browse data through normal ECS distribution methods.

Several ECS data collections are created maintained at each site to store the bulk metadata and bulk browse data. Bulk Metadata Generator and Bulk Browse Generator Tools are run daily at each site to populate the data collections. One metadata product is created per ESDT group per day. Each product contains an external representation of the metadata for each new, updated, or deleted granule that is a member of the ESDT group. The format used for the external representation of the metadata is XML. One bulk browse product is produced per day that contains references to all new, updated, or deleted browse granules. Value-added providers may use any of the standard ECS search, order, and subscription capabilities to find and order these bulk metadata and browse products.

Each of the BMGT-specific ESDTs stores products that contain an XML representation of the following kinds of data or metadata:

- ECSMETC - ECS collection-level metadata and the packaging options that may be used when ordering products from each collection.
- ECSMETG - ECS granule-level metadata.
- ECSBBR - Browse images.
- ECSMETV - ECS collection and granule valid values.
- ECSMETU - ECS granule-level QA Updates.

The ECSMETC and ECSMETG data collections store products that contain metadata for multiple collections and multiple granules. The metadata is grouped by instrument and mission except for metadata related to the MODIS instrument, which is grouped, by mission and major discipline (ocean, atmosphere, land, and snow & ice). Each product in these collections has a group identifier Product Specific Attribute (PSA) called GroupId. The mapping of specific ESDTs to groups is provided as a configuration file with the BMGT.

The ECSBBR collection stores products that contain browse product references. The ESDT has a custom acquire service that converts the browse product references into actual browse products during distribution.

The ECSMETV collection stores products that contain the entire set of valids contained within the SDSRV database for a particular instance in time.

For each of the archivable XML products, BMGT creates a .met (metadata) file. In addition, it creates a Product Delivery Record (PDR) for each type of product. BMGT places XML products, .met and PDR files in a polling directory for Ingest pick up via the Science Investigator-Led Processing System (SIPS) interface. After successful archiving, the products are pushed to ECHO via SIPS interface.

The distribution of bulk browse products is not a function of BMGT. It is managed by OMS as follows:

Normal operation of the BGMT includes the following events, which occur on a daily basis:

- The BMGT is invoked at the specified time each day via a cron job.
- The BGMT determines the date of the previous day and then executes a series of Sybase stored procedures against the SDSRV inventory database to extract metadata for all collections, granules, browse and valids that were inserted, updated, or deleted during the target day.
- For each target ESDT version that had collection level metadata inserted, updated, and/or deleted, perform the following steps:
 - **If this collection is the first collection in a group, then create a new file and append an XML representation of the packaging options to the file.**
 - **Append an XML** representation of the collection level metadata to the file.
- Insert each file, as a product, into the ECSMETC data collection setting the value for the GroupId PSA along with the starting date and ending date of the insert, update, and/or delete activity covered by this file. In this case, the starting and ending dates are the same since the period covered is a single day.
- For each target ESDT version that had granule level metadata inserted, update, and/or deleted, perform the following steps:
 - If this granule is the first granule in a group, then create a new file.
 - Append an XML representation of the metadata for each active granule to the file.
- Insert each file, as a product, into the ECSMETG data collection setting the value for the GroupId PSA along with the starting date and ending date of the insert, update, and/or delete activity covered by this file. In this case, the starting and ending dates are the same since the period covered is a single day.
- For all browse images that were inserted or deleted within a specified time period, extract the browse identifiers and associated browse file names for each browse product and insert an XML file, called the Browse Reference File (BRF) file, as a product into the ECSBBR data collection. Set the value for start and end date of the insert, update, and/or delete activity covered by this file, allowing it to be an ECS product that can be ordered and distributed via normal ECS search and order mechanisms.

- If any collections were inserted, updated, and/or deleted during the period then create a new file, append an XML representation of the valids information, and insert the file, as a product, into the ECSMETV data collection setting the starting date and ending date of the insert, update, and/or delete activity covered by this file. In this case, the starting and ending dates are the same since the period covered is a single day.

Occasionally, the cron job that automatically executes the BGMT fails to operate on a daily basis. This may happen due to a variety of reasons including software failure, hardware failure, or changes in DAAC operational priorities. When this happens, it is necessary for DAAC operations to manually invoke the tool to generate ECSMET products for all days that were missed. The sequence of operations is described in the preceding list of events.

17.10.5.1 Running the Bulk Metadata Generation Tool (BMGT)

- 1** Log in at the machine where the Bulk Metadata Generation Tool (BMGT) is installed (e.g., e4eil01 and n4eil01).
- 2** Type `cd /usr/ecs/<MODE>/CUSTOM/utilities` then press **Return/Enter**.
 - Change directory to the directory containing the Bulk Metadata Generation Tool startup script (e.g., EcBmBMGTStart).
 - <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 run the Bulk Metadata Generation Tool, at the UNIX prompt enter (as applicable):
EcBmBMGTStart
 OR
EcBmBMGTStart -P <PRODUCT>
 OR
EcBmBMGTStart -P <PRODUCT> -I <INPUTSOURCE>
-S<InputSOURCEPRODUCT>
 OR
EcBmBMGTStart -P <PRODUCT> -I <INPUTSOURCE> -L <LIST_ONLY>
-S<InputSOURCEPRODUCT>

OR
EcBmBMGTStart -P <PRODUCT> -I <INPUTSOURCE> -L <LIST_ONLY>
-S<InputSOURCEPRODUCT> -A <ARCHIVE>

OR
EcBmBMGTStart -P <PRODUCT> -I <INPUTSOURCE> -F <FILE_NAME>
-S<InputSOURCEPRODUCT>

OR
EcBmBMGTStart -P <PRODUCT> -I <INPUTSOURCE> -L <LIST_ONLY>
-S<InputSOURCEPRODUCT> -A <ARCHIVE> -F <FILE_NAME>

OR
EcBmBMGTStart -I <INPUTSOURCE> -S<InputSOURCEPRODUCT>

OR
EcBmBMGTStart -I <INPUTSOURCE> -L <LIST_ONLY>
-S<InputSOURCEPRODUCT>

OR
EcBmBMGTStart -I <INPUTSOURCE> -L <LIST_ONLY> -A <ARCHIVE>
-S<InputSOURCEPRODUCT>

OR
EcBmBMGTStart -I <INPUTSOURCE> -L <LIST_ONLY> -A <ARCHIVE> -F
<FILE_NAME> -S<InputSOURCEPRODUCT>

OR
EcBmBMGTStart -I <INPUTSOURCE> -L <LIST_ONLY> -F <FILE_NAME>
-S<InputSOURCEPRODUCT>

OR
EcBmBMGTStart -I <INPUTSOURCE> -F <FILE_NAME>
-S<InputSOURCEPRODUCT>

- <PRODUCT> specifies what product the BMGT is to generate. The value is either **Valid**s, **Collection**, **Granule**, or **Browse**.
- <INPUTSOURCE> is the input source for the processing of granules. The value is either **FILE**, or **GRANULES**.
 - **Input source can be a file (absolute path) with GranuleID's/GeoId's or both.**
 - **Input source can also be granuleId/GeoID as string as the command line option.**
 - **Multiple GranId's/GeoId's be placed inside " " as one string with spaces separating each dbID.**
- <InputSOURCEPRODUCT> is the input source product for the processing of granules. The value is either **Granule**, **Browse**, or **Both**.

- **This is the product that is generated for the list provided by INPUTSOURCE.**
- **Option “Both” will generate Granule and Browse products for the list of DbID’s specified by <INPUTSOURCE>.**
- **<PRODUCT> and <INPUTSOURCE> both are being provided as arguments. The value is either Granule, Browse, or Both and either Valid, Collection, Granule, or Browse.**
 - BMGT will generate the Product specified by <PRODUCT> along with the product specified by <InputSOURCEPRODUCT> for all ESDTs and list of dbIDs in the INPUTSOURCE.
- **<LIST_ONLY> is the output BMGT will generate. The value is either [Y/N]**
 - **If LIST_ONLY is Y, BMGT will generate XML, MET and PDR for only the list of granules in input source based on <InputSOURCEPRODUCT>.**
 - **If LIST_ONLY is Y and PRODUCT is Valid it generates XML, MET, PDR for Valid and the LIST in the input source.**
 - **If LIST_ONLY is N, BMGT will generate all the products based on other parameters including the list of granules from the Input source.**
- **<FILE_NAME> is a file name for renaming files when processing granules and Browse XML files (only).**
 - This is used to rename the Granule and Browse XML files (only) generated for the list of dbIDs from the input source.
 - **The file should have .XML/.xml extension.**
 - **If absolute path is not provided the renamed file is placed in the Output directory for BMGT.**
 - **If the renamed file cannot be placed in the directory specified (OS constraints) by the absolute path the renamed file will be placed in the Output directory instead.**
- **<ARCHIVE> is the output BMGT will generate This value can be either [Y/N].**
 - **If ARCHIVE is Y .MET and .PDR will be generated along with XML for the List of granules in Input source.**
 - **If Archive is N (this can only be when LIST_ONLY is set to Y not otherwise) .MET and .PDR will not be generated for granules in Input source.**
- **<InputSOURCEPRODUCT> is ignored if <INPUTSOURCE> is not provided.**
- **<InputSOURCEPRODUCT> is defaulted to "Both" if <INPUTSOURCE> is provided and <InputSOURCEPRODUCT> is not provided as argument.**

- If <INPUTSOURCE> is not provided BMGT ignores <InputSOURCEPRODUCT>, <LIST_ONLY>, <FILE_NAME>, and <ARCHIVE> flags.
 - If <LIST_ONLY> is N, <ARCHIVE> Flag is ignored.
 - The BMGT produces ECSMETC, ECSMETG, ECSBBR and ECSMETV products, which consist of XML files.
 - The BMGT runs and the BMGT ALOG file, EcBmBMGT.ALOG, and debug log, EcBmBMGTDebug.log, record errors, warnings, and information about utility events.
-

17.10.6 Running the Bulk URL Utility

The **Bulk URL Utility (EcBmBulkURL)** exports to ECHO the metadata content of products in the Data Pool. It can be run with either an “Insert” option or a “Delete” option.

With the “Insert” option, the Bulk URL utility is run on a daily cron job or from the command line, and should be run after the daily BMGT run has completed (this maximizes the probability that ECHO has already received information about the granules from the BMGT and has added the granules to its inventory).

With the “Delete” option the Bulk URL utility is run as an integral part of the Data Pool Cleanup utility. When the Cleanup utility is run, the operator chooses whether to run the cleanup as a one pass (setting the echo mode command line parameter to “deleteall”) or two-pass (setting the echo mode command line parameter to “predelete” on the first pass, and setting the echomode command line parameter to “finishdelete” on the second pass) process.

The procedure for running the Bulk URL Utility is based on the following assumptions:

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

17.10.6.1 Running the Bulk URL Utility

- 1 Log in at the host for the Bulk URL Utility (e.g., x4dp101).
- 2 Type **cd /usr/ecs/<MODE>/CUSTOM/utilities** then press the **Return/Enter** key.
 - The working directory is changed to cd /usr/ecs/<MODE>/CUSTOM/ utilities directory which contains the Bulk URL Utility start script (.g., EcBmBulkURLStart).
- 3 Type **EcBmBulkURLStart <MODE> [Insert | Delete]** then press the **Return/Enter** key.
 - <MODE> is the mode in which the utility is being executed (e.g., OPS, TS1, or TS2).
 - **Insert** option - the Bulk URL utility will export an xml representation of the ftp URL information for science files, metadata files, and browse files associated with Data Pool granules, which meet all of the following conditions:

- The granule is in a collection that is eligible for export (from Data Pool) to ECHO.
 - This is determined by whether the bulkExportedFlag = “Y” for that collection in the DICollections table in the Data Pool database. (NOTE: The BMGT determines which collections are eligible for export (from SDSRV) using the EcBmBMGTGroup.xml file in the /usr/ecs/<mode>/CUSTOM/cfg directory on x4eil01, but the Bulk URL utility uses the bulkExportedFlag in DICollections instead.) DAAC operations staff should set the bulkExportedFlag for each collection that is also exported by the BMGT, according to DAAC export policy. The bulkExportedFlag should then only be changed if the export policy changes.
- The granule was inserted into the Data Pool during the time period specified by the run. This time period is either:
 - A time period ending when the Bulk URL utility begins its run, if the EcBmBulkURLConfigParams.xml file has doPreviousFlag = true; the time period can be in days (duration = day and count = <number of days>), or hours (duration = hour, and count = <number of hours>); or
 - The time period between a specified start and end date/time, if doPreviousFlag = false and <start Date> and <end Date> are specified in the EcBmBulkURLConfigParams.xml file.
- The granule is not an “order only” granule, i.e., it was not placed in the Data Pool for the sole purpose of fulfilling a specific OMS order. If the granule is an “order only” granule, the isOrderOnly flag in the DIGranules table will have been turned on when the granule was inserted in the Data Pool.
 - This constraint was introduced with Release 7.0, when the Data Pool began to be used as a staging area for OMS orders. The intent of the Bulk URL utility is only to export Data Pool URLs for granules placed in the Data Pool by subscription or batch insert – these granules are intended to be available to the general Data Pool user community.
- **Delete** option - the Bulk URL utility exports an XML representation of a list of granuleURs for Data Pool granules, which meet all of the following conditions:
 - The granule is in a collection that is eligible for export (from Data Pool) to ECHO.
 - This is determined by whether the bulkExportedFlag = “Y” for that collection in the DICollections table in the Data Pool database.

- The granule will be deleted from the Data Pool in this (possibly two-part) run of the Cleanup utility (i.e., the granule is represented in the DIFilesToBeDeleted table in the Data Pool database).
 - In the one-pass process (echomode = “deleteall”), Cleanup deletes expired granules from the Data Pool and invokes the Bulk URL utility to notify ECHO of the deletions. If the one pass Cleanup process is used, it is possible for a granule to be deleted from the Data Pool before ECHO removes the granule from its inventory, thus making it possible for an ECHO user to receive an error when attempting to access an already-deleted granule.
 - In the two-pass process, during the first pass (echomode = “predelete”), Cleanup determines which granules are expired, invokes the Bulk URL utility to notify ECHO of their pending deletion, but does not actually delete the granules. After a time lag during which ECHO removes the to-be-deleted granules from its inventory, Cleanup is run in the second pass (echomode = “finishdelete”) and all of the to-be-deleted granules are actually removed from the Data Pool. The Bulk URL utility is not invoked during the second pass of the Cleanup utility.
 - The granule is not “order only.”
 - For further information concerning the Data Pool Cleanup Utility refer to the procedure for **Running the Data Pool Cleanup Utility** (previous section of this lesson).
 - The Bulk URL Utility runs and the ALOG file, EcBmBulkURL.ALOG, and debug log, EcBmBulkURLDebug.log, record errors, warnings, and information about utility events.
 - The Bulk URL Utility does not perform automatic recovery operations. If the log file indicates that the utility failed to run to successful completion, the utility should be reinvoked.
-

17.10.7 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.

17.10.7.1 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 `DPrecentInserts_<YYYYMMDD>`, is located at the top-level Data Pool directory. It contains distinct GroupID, ShortName, and VersionID. For example, the file `DPrecentInserts_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      MOD15A2     077
MOAT      MYD01       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 `DPRRecentInserts_<ShortName>_<VersionID>_<YYYYMMDD>`. The files contain `SHORTNAME`, `VERSION_ID`, and `RELATIVE_PATH`. For example, the file `DPRRecentInserts_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.

17.10.8 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 exists 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.

17.10.8.1 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** 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., x4dp101).
- 3 Type `cd /usr/ecs/<MODE>/CUSTOM/utilities`, then press the **Return/Enter** key.
- 4 Type `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.
 - For example:
EcDIRemap.pl OPS –esdt MOD29 –version 4 –oldgrp MOST –newgrp MOSS
 would remap collection MOD29, Version 4, (i.e., MOD29.004) from collection group MOST to collection group 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.

17.10.9 Running the Data Pool QA Update Utility

The **QA Update Utility (EcDIQaUpdate)** provides the DAAC Operations Staff with a command-line interface for updating the QA data for granules in the Data Pool inventory. Both the inventory and the corresponding XML files on disk are updated.

The QA Update Utility can operate in either of two modes, depending on the command-line parameters:

- Single file – the utility reads an input file specified on the command line.
- Multiple files – the utility reads multiple files from the QAMUT’s undo directory (e.g., /usr/ecs/OPS/CUSTOM/data/DSS/QAMUT/QAMUTUndo).

Each input file contains a list of granule db ids and the QA updates to be performed for those granules. The files, called “undo” files, are generated by the QAMUT utility in its “undo” directory. Therefore, the QA Update must be run after QAMUT has been run.

The procedure for running the QA Update Utility is based on the following assumptions:

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

17.10.9.1 Running the QA Update Utility

NOTE: The QA Update Utility is typically run after the QAMUT utility has run because output from the QAMUT is used as input to the QA Update Utility.

- 1** Log in at the host where the **QA Update Utility** is installed (e.g., **x40ml01**).
- 2** Type **cd /usr/ecs/<MODE>/CUSTOM/utilities** and then press the **Return/Enter** key.
- 3** Type **EcDIQaUpdateStart -mode <MODE> [-file <filename>]** and then press the **Return/Enter** key.
 - <MODE> is the mode in which the utility is being executed (e.g., OPS, TS1, or TS2).
 - <filename> is the name of a specific file to be used as input to the utility. The file must reside in the QAMUT undo directory (e.g. /usr/ecs/OPS/CUSTOM/data/DSS/QAMUT/QAMUTUndo). The -file parameter is optional. If no -file parameter is provided, the utility uses all undo files in the QAMUT undo directory as input.
 - The QA Update Utility runs and the log file, EcDIQaUpdate.log, records errors, warnings, and information about utility events.

- The QA Update Utility is able to recover from an execution failure caused by situations such as system faults or database errors leaving all or some of the QA updates unprocessed. The utility detects such failures during the next run and continues processing the QA updates that were left unprocessed in the previous run. The operator is given no choice as to recovery: either recovery proceeds or the Data Pool inventory and disk files are corrupted.
-

17.10.10 Running the Data Pool Move Collections Utility

The **Data Pool Move Collections Utility (EcDIMoveCollection)** provides the DAAC Operations Staff with 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 (destination), followed by removal of the old collection directory (origination) and its contents. The utility then establishes a link to the new location in place of the old directory. As a result, existing URLs are not invalidated by the move and no updated URLs need to be exported to ECHO. However, any existing URLs and file pointers become invalid from the time when the utility starts deleting the existing directories until the time the link is established. During that time the following errors could be encountered with respect to the collection being moved:

- A Data Pool ftp user or an EDG user clicking on a URL might experience a temporary error when trying to access files and directories. File transfers already in progress at the beginning of the deletion should complete normally.
- FTP Pull users could experience similar temporary problems trying to access links in FTP Pull directories that were established by the OMS and that point to granules in the moving collection.
- The Data Pool Web GUI returns an error if a user tries to access the collection via a bookmark. It flags the collection and does not display it as an active link on the collection drill down web page, thus temporarily preventing drill down access to the collection.
- The Data Pool insert service looks up the collection path in the Data Pool database during the insert process. The collection path is updated once the copy step is complete. Any Data Pool insert processes that looked up the copy path before it was updated insert their granules into the old directory location. If those granules are not copied but are then removed, they become phantoms and could trigger additional errors downstream, e.g., in distribution. Alternatively, if they were not removed, they would cause the move process to fail, as the Data Pool insert service would re-create the deleted directories.
- 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 fails and causes an operator intervention. Since the time

window between file location look up and ftp push start is small, the chances for that occurring are very small. In such cases the operator would need to resubmit the request, and since the directory entry would now have been updated, the ftp push operation would succeed.

If the preceding 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.

The procedure for running the Data Pool Move Collections Utility is based on the following assumptions:

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

17.10.1 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 Type `cd /usr/ecs/<MODE>/CUSTOM/utilities`, then press the **Return/Enter** key.
- 3 Type `EcDlMoveCollection.pl <MODE> -shortname <shortname> -versionid <versionID> -targetfs <path> [-verbose]` and then press the **Return/Enter** key.
 - *<MODE>* is the mode in which the utility is being executed (e.g., OPS, TS1, or TS2).

NOTE: When the utility is run in OPS mode an operator prompt is displayed to prevent any accidental loss of data.

- *<shortname>* is the shortname of the collection being moved. It is a mandatory parameter.
- *<versionID>* is the version identifier of the collection being moved. Leading zeros must not be included. `-versionid` is a mandatory parameter.
- *<path>* is the relative path to the target file system (the file system 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.
- `-verbose` is an optional parameter. When `-verbose` is specified, some information is displayed on the screen and detailed information is written to the utility's log. Nonverbose is the default.
- For example:
EcDlMoveCollection.pl OPS -shortname MODVOLC -versionid 1 -targetfs fs1 -verbose would move collection MODVOLC, Version 1, (i.e., MODVOLC.001) from its current directory as specified in the database to the new filesystem fs1. The collection would be moved from /datapool/OPS/user/MOAT to

/datapool/OPS/user/fs1/MOAT. The utility would run using the verbose option, which would display information to screen and to the log.

- The Data Pool Move Collections Utility runs and the log file, EcDIMoveCollection.log, records errors, warnings, and information about utility events.
 - The Data Pool Move Collections Utility is from an execution failure caused by situations such as system faults or database errors leaving all or some of the file moves unprocessed.
 - At startup the utility determines whether or not an execution failure has occurred. If there has been an execution failure, the utility prompts the operator as to whether or not to attempt recovery. If the operator chooses to recover, the utility completes the processing of file moves that were left unprocessed in the previous run. Upon completion of the recovery, the utility runs again with the current command-line parameters.
 - **An operator** may not wish to recover (e.g. if the target filesystem has become corrupted or full). In this case, recovery is not attempted, and the utility runs with the current command-line parameters. The moveFlag is automatically reset to “N”. Any files that were copied to the target file system that experienced the failure would have to be deleted manually.
-

17.10.11 Running the Data Pool Hidden Scrambler Utility in Rename Mode

In Synergy V a new capability was introduced whereby 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 pre-Synergy V directory naming conventions so they can deal with granules in the hidden directory structure.

In Synergy V, 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 FtpPull 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.).

WARNING

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.

17.10.11.1 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:

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.

17.10.12 Data Pool Non-ECS Action Driver Insert Tool

The **Non-ECS Action Driver (EcDINdActionDriver)** tool provides the DAAC Operations Staff with a command-line interface for inserting Non ECS granules into the Data Pool. The command is found in the /usr/ecs/OPS/CUSTOM/bin/DPL directory. This tool runs only if the Data Pool database server is running and if the database is available.

The procedure for running the **Non-ECS Action Driver** is based on the following assumptions:

- Data Pool database server is running.
- Non-ecs data has been queued for insertion using the Batch Insert tool (see section 17.11.4).
- Data Pool database is available.

17.10.12.1 Running the Non-ECS Action Driver

- 1 Log in at the host where the **Non-ECS Action Driver** is installed (e.g., x4dpl01).
- 2 Type `cd /usr/ecs/<MODE>/CUSTOM/bin/DPL` and then press the **Return/Enter** key.
- 3 Type `EcDINdActionDriver <MODE> [configFile <filename>]` and then press the **Return/Enter** key.
 - <MODE> is the mode in which the utility is being executed (e.g., OPS, TS1, or TS2).
 - <filename> is the name of a specific configuration file to be used as input to the utility. This field is optional. The default is /usr/ecs/<MODE>/CUSTOM/cfg/EcDINonECSActionDriverCFG.

- The Non-ECS ActionDriver tool queues Non-ECS data for insertion in the Data Pool. Insertions may be viewed using DPL Maintenance GUI **List Insert Queue** tab.
-

This page intentionally left blank.