

NASA/TM—2005–209997/VER200404.0/VOL. 6a



ICESat (GLAS) Science Computing Facility Document Series

Volume 6a SCF Interface Software Operator's Guide for rSCF Version 200404.0

*Kristine Barbieri
Anita Brenner*

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

April 2004

TABLE OF CONTENTS

TABLE OF CONTENTS	II
SECTION 1: INTRODUCTION	1
SECTION 2: RELATED DOCUMENTATION	2
SECTION 3: OVERVIEW	3
SECTION 4: DIRECTORY STRUCTURE	4
SECTION 5: DATA.....	5
SECTION 6: PDR CREATION	6
SECTION 7: DATA TRANSFER CRON JOB.....	7
SECTION 8: OPERATION.....	8
SECTION 9: TROUBLESHOOTING GUIDE.....	9
SECTION 10: ERROR EMAIL RECIPIENTS.....	11
APPENDIX A: ABBREVIATIONS & ACRONYMS	12

SECTION 1: INTRODUCTION

A series of Perl scripts have been written to move data from the GLAS Main Science Computing Facility (mSCF) to the Remote Science Computing Facilities (rSCFs). This document gives an overview of how the scripts are run and troubleshooting techniques for when there are problems.

SECTION 2: RELATED DOCUMENTATION

- *SCF Architectural Design Document*
- *SCF Data Request Software Detailed Design Document*
- *SCF Data Request Software User's Guide*
- *SCF Data Visualization Software User's Guide*
- *SCF Operator's Guide*
- *SCF Interface Control Document*
- *SCF Interface Software Update Document*
- *SCF Interface Software Installation Guide*
- *SCF Interface Software Detailed Design Document*
- *SCF Interface Software Operator's Guide for rSCF*
- *SCF Interface Software Operator's Guide for mSCF*
- *I-SIPS Interface Software Operator's Guide for ISF*
- *I-SIPS Interface Software Operator's Guide for UTCSR*
- *Interface Control Document Between I-SIPS/ISF and CSR*

SECTION 3: OVERVIEW

All data transferred from the mSCF to the rSCF will be put in an rSCF output cache on the mSCF to be pushed by the mSCF. A Product Delivery Record (PDR) will be placed in the cache listing the files contained in the distribution along with validation information. Once the files and PDR are placed in the output cache, a PDR transfer file (XFR) is created denoting that the files and PDR are ready to be transferred. After the files have been pushed by the mSCF to the rSCF input cache, the rSCF will place a Product Acceptance Notice (PAN) and its associated XFR into the rSCF input cache on the mSCF. The PAN will indicate whether the files were successfully transferred or if an error occurred. If the files were successfully transferred to the rSCF, then the mSCF will remove the files from its output cache. If the file transfer was unsuccessful, email will be sent to the mSCF explaining the error in detail. Once the data is in the input cache on the rSCF, the data will be moved to an appropriate subdirectory based upon information in the PDR. Once the files in the subdirectory are validated against the PDR, the PDR and XFR will be removed from the input cache.

Likewise, all data transferred from the rSCF to the mSCF, along with a PDR and XFR, will be put in an output cache on the rSCF and be pushed by the rSCF to the input cache on the mSCF. Once the push is complete, the mSCF will put a PAN and XFR in the input cache on the rSCF indicating a successful transfer or not. If the files were successfully transferred to the mSCF, then the rSCF will remove the PAN and the files from its output cache. If the file transfer was unsuccessful, email will be sent to the rSCF explaining the error in detail.

SECTION 4: DIRECTORY STRUCTURE

Data to be pushed to a rSCF must be placed in that rSCF's output cache on the mSCF. The data is then pushed to the input cache on the rSCF. Likewise, the rSCF will push data from its output cache to that rSCF's input cache on the mSCF. Below is a table indicating the directory names of the input and output caches on the mSCF and the rSCFs. These directory names utilize the site names explained in the *SCF Interface Software Installation Guide*.

Host	Input Cache	Output Cache
mSCF	/SCF/ingest/ALT	/SCF/dist/ALT
mSCF	/SCF/ingest/LIDAR	/SCF/dist/LIDAR
mSCF	/SCF/ingest/MIT	/SCF/dist/MIT
mSCF	/SCF/ingest/NSIDC	/SCF/dist/NSIDC
mSCF	/SCF/ingest/OSU	/SCF/dist/OSU
mSCF	/SCF/ingest/UCSD	/SCF/dist/UCSD
mSCF	/SCF/ingest/UTCSR	/SCF/dist/UTCSR
mSCF	/SCF/ingest/UW	/SCF/dist/UW
mSCF	/SCF/ingest/WFF	/SCF/dist/WFF
rSCFs	/SCF/ingest	/SCF/dist

SECTION 5: DATA

Data created by the mSCF and transferred to the rSCF includes:

- Any portion of products GLA01-GLA15 as requested via subscription or special request
- Browse products for each product file
- One pass table for each product file
- One unique record index table for each product file
- One or more bin tables
- A georeference table for each bin table
- Updated rev file

Data created by the rSCF and transferred to the mSCF includes:

- Science team QA
- Level 3 and 4 Products

SECTION 6: PDR CREATION

Once a distribution set has been created at a rSCF for transfer to the mSCF, a Product Delivery Record (PDR) must be created to accompany it. This PDR contains benchmark information about the data including number of files, file sizes, and file checksums so that the data may be verified after transfer.

To create a PDR, a PDR Input File must be created and the script, `create_pdr.pl`, must be invoked. Refer to the *SCF Interface Software Detailed Design Document* for a description of the script and for the PDR Input File format. The PDR Input File and PDR must be created in the directory where the data is originally created since the data may become corrupt during first move or copy. Once the PDR Input File has been created, its name should be put as the input argument to the script in the file `/SCF/src/perl/ops/run_create_pdr.ksh`. The script then must be invoked with the command:

```
/SCF/src/perl/ops/run_create_pdr.ksh
```

The script creates a PDR with the file name "SITE_NAME.DISTRIBUTION_ID.PDR". The determination of the DISTRIBUTION_ID is TBD. The site name is an environmental variable in the `run_create_pdr.ksh` file and indicates the GLAS site that the PDR is created at. Valid site names are listed in the "Directory Structure" section. This script also generates a PDR Transfer File (XFR) with the name "PDR_filename.XFR" denoting the completion of the PDR file.

Once the PDR is created, the distribution set, PDR, and XFR must be moved, copied, or linked to the output cache, so that the data may be pushed to the mSCF via the cron job (explained in the next section).

SECTION 7: DATA TRANSFER CRON JOB

The data transfer scripts to move data are invoked automatically at set time intervals via a cron job on the rSCF systems. Access to the cron job can only be done under the "scf" account. To view the cron file which has the times of invocation type *crontab -l*. If the cron times need to be changed, refer to the crontab man page on how to setup a cron job. Below is an example of how a cron job is setup:

- To edit the cron file in vi: *crontab -e*
- Type the following line in the file to invoke the run_scripts.ksh script every 10 minutes starting at 5 minutes after the hour, every hour, everyday at the rSCFs:

```
05,15,25,35,45,55 * * * * * /SCF/src/perl/ops/run_scripts.ksh
```

Currently, the data transfer cron jobs are set to run once per minute:

```
* * * * * /SCF/src/perl/ops/run_scripts.ksh
```

- Write and quit out of the file to invoke
- To view the cron file: *crontab -l*
- To remove the cron file: *crontab -r*

All output will be appended to the files /SCF/tmp/cron.log.###. There are 10 log files and they are overwritten in turn as they reach their maximum size limit. A grep on the date may be done to determine which file contains information for that date.

The cron invokes all scripts called from the main script, run_scripts.ksh. On the rSCF systems, the following scripts are performed when the cron job is invoked:

- files_to_subdir_from_push.pl: Files are moved from the input cache to a subdirectory and send a PAN back to the mSCF

SECTION 8: OPERATION

Since the scripts run automatically via the cron job, the operator only needs to ensure that incoming files have been received and outgoing files have been received and deleted from the output cache.

No file will be transferred until there is a PDR for the file and an XFR present to ensure that the PDR file is complete.

When files are pushed by the mSCF, they are first put in the input cache then moved to a subdirectory. Below is a table showing the subdirectory where the different files received from the mSCF should be expected:

Data Received from mSCF	Expected Subdirectory
Any portion of products GLA01-GLA15	/SCF/product_sets/xxxx
Browse products	/SCF/product_sets/xxxx
Pass tables	/SCF/product_sets/xxxx
Unique record index tables	/SCF/product_sets/xxxx
Bin tables	/SCF/product_sets/xxxx
Georeference tables	/SCF/product_sets/xxxx
Rev file	/SCF/ancillary_data/rev_files

where xxxx designates the subdirectory associated with the \$SUBDIR keyword in the PDR file. If this is missing, the files will go directly into /SCF/product_sets/default.

A log file is included with each batch of files delivered to the rSCF. It is put into /SCF/product_sets/logs. The name of the log file includes the name of the user and the date the file is created. It contains the parameters of the request and any errors that may have occurred during processing. A log file will be received for a request even if no files are received and will indicate why no files were processed (i.e. no data were found matching the request parameters).

When files are pushed to the mSCF, the XFR is removed from the output cache, but the PDR and data files remain. Sometime later, a PAN should be received in the input cache. A script checks the PAN for the keyword "Successful" and if present, removes the PAN from the input cache and the corresponding PDR and all of the files listed in it from the output cache. Therefore, if the files and PDR are gone from the output cache, it can be assumed that the data transfer to the mSCF was successful.

The run_scripts.ksh script invoked by the cron job appends or overwrites a file, /SCF/cron.log.### where ### is 001 to 010, every time it runs. Starting with cron.log.001, the file is appended until it reaches a certain size, then the next file, cron.log.002, is written. If the file does not exist it is created, otherwise it is overwritten. In this way, there is no need for cleanup since the log files just overwrite themselves and since there are ten files, error messages are kept for a while in case problem troubleshooting is necessary. The number of log files and max size of each file may be modified as necessary in the run_scripts.ksh script.

SECTION 9: TROUBLESHOOTING GUIDE

If an error occurs during a data transfer, the PAN will indicate a problem, but an email will be sent explaining the problem in further detail. The cron.log.### file may also show a problem. Use the error description in the email to troubleshoot the problem. Once the problem has been solved, all the files must be put into the output cache again so that they will be transferred when the cron job invokes. Since the scripts reference the complete PDR, it is important that all of the files listed in the PDR are available to be transferred again, even if some of them were already transferred before the transfer was interrupted. In most cases, the XFR file may have been removed, since this is the first thing that most of the scripts do. The scripts read the PDR or PAN only if the XFR is also present, and therefore remove the XFR first so that another cron job does not try to access the same PDR or PAN at the same time. Therefore, if a transfer has to be run again, if the XFR is missing, it needs to be recreated. This can be accomplished simply by using the "touch" command since the XFR only needs to exist, not necessarily contain anything. The XFR file name is simply the PDR or PAN file name with the extension ".XFR" added at the end. For example: "touch filename.PDR.XFR".

When files are moved from the input cache to the subdirectory, they are actually first copied to a .tmp directory under the subdirectory. The .tmp directory has the request ID or distribution ID appended to the name to ensure a unique directory name. When the file sizes have been verified against the PDR they are then moved to the subdirectory. This is to ensure that the files are not accessed before the move is complete. If a problem occurs during the copy (i.e. the subdirectory runs out of space), the files and PDR still reside in the input cache. Once the problem is resolved, just "touch" the XFR to attempt the move again. Sometimes this subsequent attempt will fail if the permissions of the files in the .tmp directory do not allow the files to be overwritten. This may happen when a copy is abruptly cut off. If need be, go to the subdirectory listed in the PDR and type "ls -la .tmp*" to see the directories listed. Type "rm -rf .tmp*" to remove them. Retouch the XFR in the input cache and the move should work.

Below are common error messages generated by the Perl scripts and a general outline of troubleshooting techniques if an error occurs. Refer to the *SCF Interface Software Detailed Design Document* for detailed descriptions of the PDR, XFR, and PAN files.

Error Type	Check that:
Error creating file	<ul style="list-style-type: none"> - Disk space is available * - Permissions allow write access **
File does not exist	<ul style="list-style-type: none"> - Full file name is correct - Permissions allow read access **
Error opening file	<ul style="list-style-type: none"> - Permissions allow read access ** - Full file name is correct
Cannot create directory	<ul style="list-style-type: none"> - Disk space is available * - Permissions allow write access **
Number of files has not been read from PDR	<ul style="list-style-type: none"> - PDR follows correct format
Number of files does not match number in PDR	<ul style="list-style-type: none"> - PDR follows correct format - Disk space is available * - All files listed in PDR exist in the directory - File transfer was complete
PDR is for a subscription and number of files is 0	<ul style="list-style-type: none"> - PDR follows correct format - PDR is really for a subscription
Checksum for file does not match checksum in PDR	<ul style="list-style-type: none"> - PDR follows correct format - File checksum is accurate - File transfer was complete

PDR verification did not pass all tests	<ul style="list-style-type: none"> - PDR follows correct format - All files listed in PDR exist in the directory - File checksums are accurate - File transfer was complete
PAN file does not indicate successful transfer	- Look for accompanying email for error details
Expected data can not be found	- Cron job ran

* **If disk space is not available:** Have system administrator create more disk space on system

** **If file has restrictive permissions:** Have system administrator change permissions on file

SECTION 10: ERROR EMAIL RECIPIENTS

If an error occurs during a data transfer an email is sent explaining the problem. At the mSCF, a system alias has been created to direct email to mSCF personnel:

send_mail_mscf@icesat0.gsfc.nasa.gov

This alias is used by the following routine:

`/SCF/src/perl/ops/send_mail_local.pm`

APPENDIX A: ABBREVIATIONS & ACRONYMS

GLAS	Geoscience Laser Altimeter System
GSFC	Goddard Space Flight Center
ICESat	Ice, Cloud, and land Elevation Satellite
ISF	Instrument Support Facility
I-SIPS	ICESat Science Investigator-led Processing System
LIDAR	LIght Detection And Ranging
MIT	Massachusetts Institute of Technology
mSCF	Main Science Computing Facility
NASA	National Aeronautics and Space Administration
OSU	Ohio State University
rSCF	Remote Science Computing Facility
UCSD	University of California at San Diego
UTCSR	University of Texas Center for Space Research
UW	University of Washington
WFF	Wallops Flight Facility