Basic Data Handling Guide
CATSCAN Experiment
2004.031

This is a general guide to handling the CAT/SCAN data after the CompactFlash disks or memory cards have been collected from the sites in the field. The first portion of this can be done in the field to determine how well a site is performing before leaving the site.

MOVE DATA FROM THE CARDS TO THE LAPTOP

1. **Start the Linux laptop.** Two choices for boot kernel will be displayed. Select the first one which should be the one that **does not** end in "-8". The computer should finish booting at the desktop and already logged into the "field" account.

2. **Start RACS.** From the desktop click on the laptop-shaped icon in the row of icons along the bottom of the display.


3. **Enable the RAID system.** A window displaying three red dots should appear. You may need to move the window up so you can see the buttons at the bottom. Click on the **Enable** button. The first red dot, labeled Archive, should turn green after about 1 minute or less.
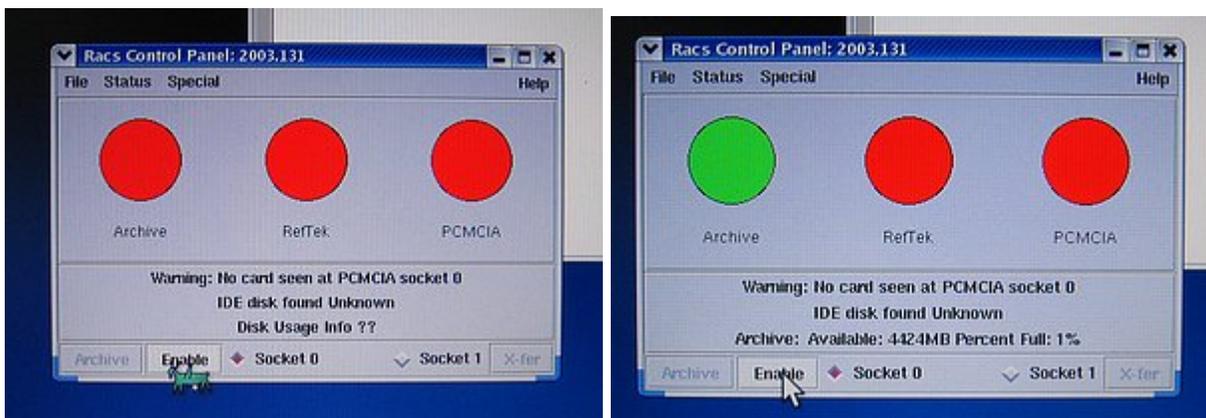
4. **Insert** one of the CompactFlash disks or memory cards into the PC Card Adapter and slide it into the upper PC Card slot on the right side of the laptop. **You should hear two beeps** from the laptop. If you do not hear two beeps or if you hear a series of more than two beeps eject the card and try again. If after multiple tries you still do not hear two beeps then reboot the laptop and start over from step 1. This problem may occur after several cards have been read during one session, but usually just ejecting the card and reinserting it fixes the problem.

5. **Enable the card.** Once the above step has been completed click the Enable button again. The remaining two red dots, labeled Reftek and PCMCIA, should turn green after a few moments. The Enable button will change to the Swap button.
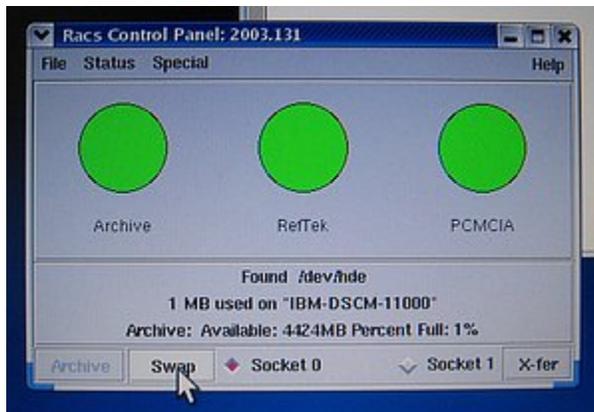
6. **Set the output prefix.** Pull down the File menu and select Output prefix... A dialog box will appear. Enter the station code into the text field and click OK. The code that you enter, plus the day-of-year number that the system will add, will be the name of the directory that will be created and where the raw data file read from the memory card will be placed. This directory will always be created in the /data directory. The name following the label "Output prefix:" will usually be whatever was last used (possibly the last station downloaded). There are times when this default value is blank. If nothing is entered the system will create a directory named ".<day number>", which will not show up with an ls command since a name beginning with a "." is hidden. If you cannot find your data check for a hidden directory using ls -a.
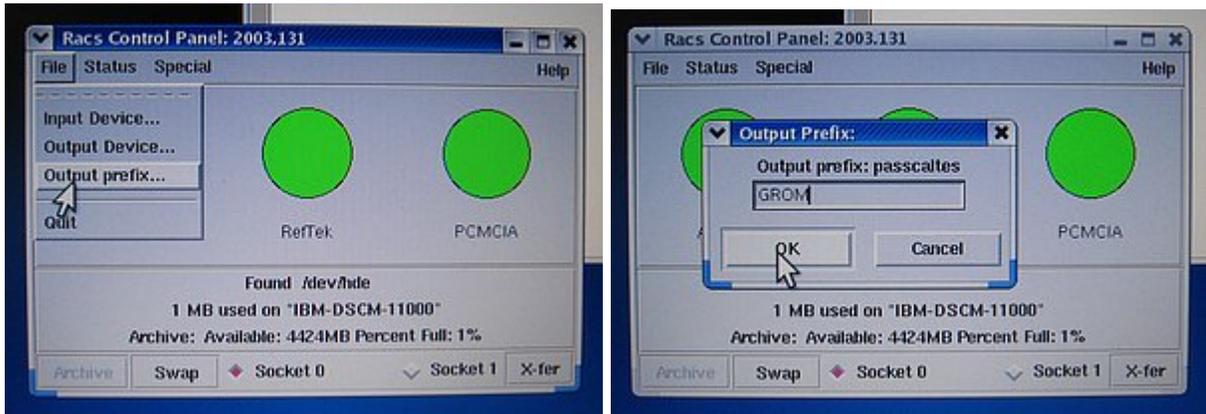

7. **Transfer the data.** Click on the X-fer button and the transfer should begin. During the transfer the DAS serial number will be displayed. You should check to make sure that the DAS serial number matches the station that the card was taken from. The green Archive dot will flash while the data is transferred. It may take up to 20 minutes to read all of the data from a 1GB card. The transfer rate is about 1MB/sec plus processing time to create the raw data file.

8. **Check the transfer results.** When the transfer has been completed change directory to /data and check that the correct directory name was created, and that the ".ref" raw data file was created in that directory. The file name of the raw data file will contain the start time of the data (when acquisition was started) and the DAS serial number.
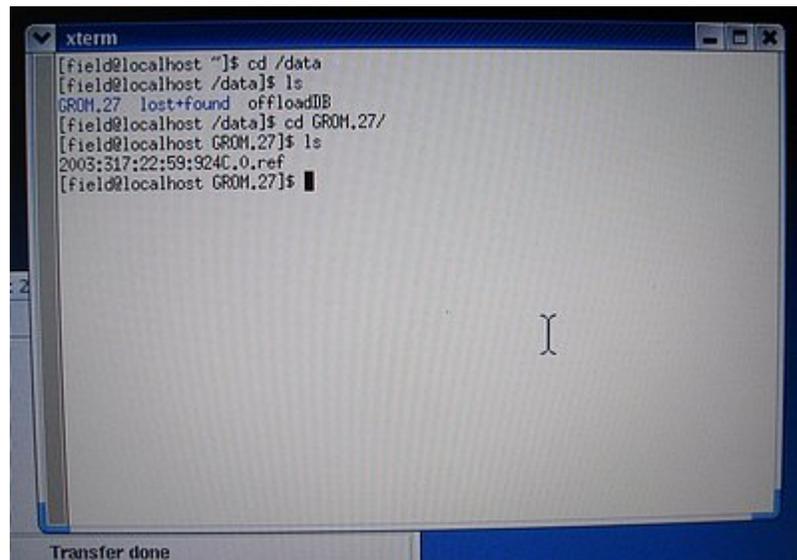
9. **Swap out the card.** Click on the Swap button on the RACS control panel. This will disable the card, and the PC Card slot. The Reftek and PCMCIA dots will turn red, and the Swap button will turn back into the Enable button. After the dots have turned red you may remove the memory card from the PC Card slot of the laptop. Go back to step 4 if you have more memory cards to read.
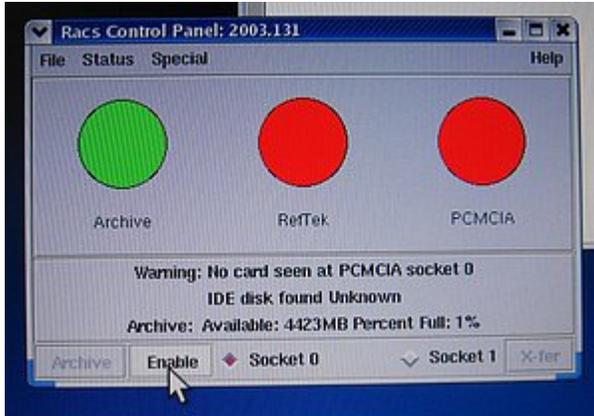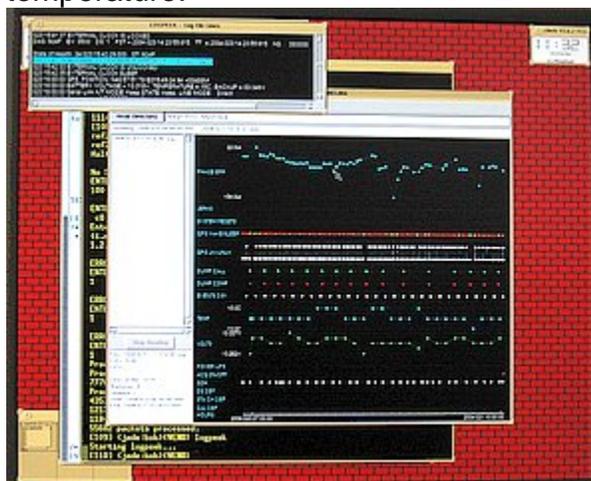

10. **Look at the data.** At this point you can process the data to SEGY using the program ref2segy and a command like

    ref2segy -f <.ref file name>

This will create files ending in .log, .run, possibly .err, and "R" sub-directories. There will be an "R" sub-directory (such as "R031.01") for each day that the time range of the .ref file covers. A CRITICAL error message will be displayed when the program starts which can be ignored. In some cases ref2segy may require information about the project that was not written to the raw data. It may ask for the sample rate of Data Stream 1 (enter **50** or **100**, depending on what the sample rate was during the time period covered by the raw data file), for the data format (enter **c0** (cee-zero), which stands for compressed), for the list of channels for data stream 1 (enter **1,2,3**), and for the gain for each channel (enter the number **1** for each channel).

Run the program logpeek (just type logpeek) to produce a graphical representation of the contents of the .log file. This will plot information such as timing information, GPS status items, battery voltages and DAS temperature.

Examine the trace data by changing directory to one of the "R" directories and using the program pql

    pql <something>

<something> can be any combination of standard wildcards to examine any of the files you want to see. The files in the "R" directories ending in .1 are the trace files for the vertical channel, .2 for the N-S channel, and .3 for the E-W channel. There should be a trace file for each channel for each hour of the day.

11. **Rename and cleanup.** Rename the directories created by the RACS program from, for example, GROM.27 to just GROM. This will make things easier during later processing, and this is the convention that has been used on the Sun workstation. Change directory to the /data directory and run the script cleanup.sh. This will go through all of the directories and remove files created by the ref2segy program (you won't want them taking up space on the tape when the data is backed up -- they can easily be recreated).

MOVING DATA TO THE SUN WORKSTATION

12. **Keep the RACS control panel running.** To be able to "see" the /data directory the laptop the RACS control panel program must be kept running, and the Archive dot must be green (Archive enabled).

13. **Start up the Sun workstation.** First power up the DLT drive using the power switch on the back of the unit. Start up the external disk drive with the blue power button on the front. Turn on the monitor. Start the CPU with the power button on the front. The computer will either boot all of the way to the graphical login prompt, or you will have to type "boot" at the OK prompt.

14. **Login.** When the workstation finishes booting up log in to the field account.

15. **Connect the Ethernet cable.** If necessary, connect the Ethernet cross-over cable between the laptop and the workstation. If the Ethernet has been installed then check that both computers have been connected to the building network cabling.

16. **Start gFTP.** Go to the Red Hat icon on the laptop and start gFTP which should be in the menu Internet|More Internet Applications|gFTP. The program is fairly straightforward to use. Transfer the laptop directories, /data/<station>, created by the RACS system to the workstation directories /data/<station> (the external disk drive). The program will check to see if files on the laptop already exist on the workstation, so multiple copies of the files can be kept without then wasting time by being transferred multiple times. The file structure on the workstation should be kept the same as that on the laptop, that is, in the /data directory on the laptop, and in the /data directory on the workstation there should be station directories, such as GROM, SGIO, CAMP, with the raw data ".ref" files for each station in their respective directory.
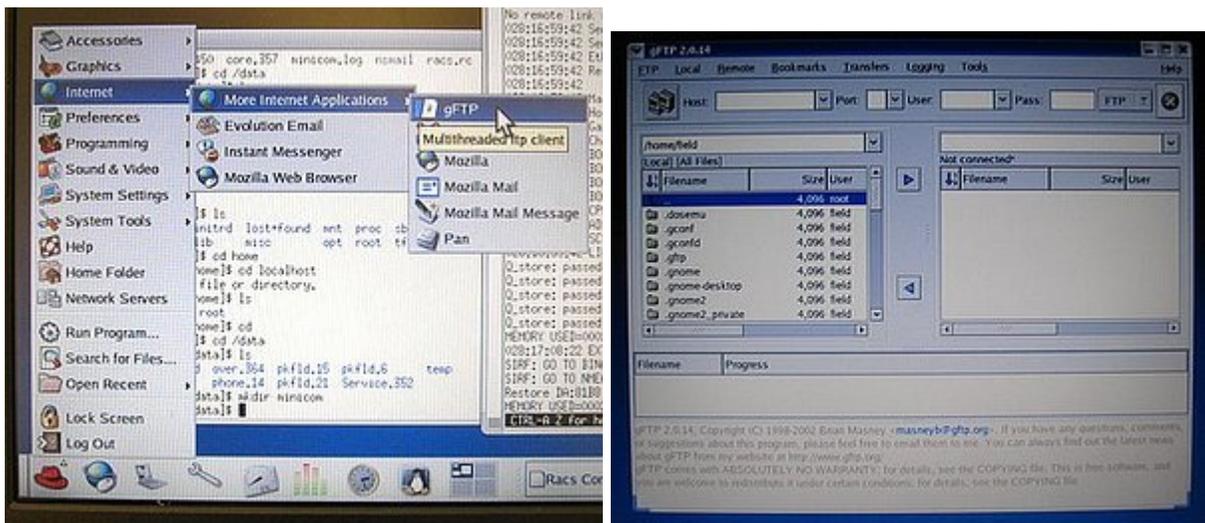
BACKING UP THE DATA TO TAPE

17. **Insert a clean tape.** Flip the lever up on the DLT drive below the slot where the tape slides into, insert the tape, and flip the lever back down. Wait until the tape has finished loading, which can take a while. To use the scripts for backing up the tape **MUST BE new, have been erased by degaussing, or have been erased** with a command like

    mt -f /dev/rmt/0 erase

Erasing a tape with mt requires about two hours.

You may APPEND data to a tape that had data previously copied to it using the backup.sh script, but if your intention is to start over from the beginning of the tape then the tape **MUST HAVE** been erased as described above first.

18. **Edit backup.sh.** In the /data directory there should be the file backup.sh. This script contains lines that run the script tar.sh. The inputs to tar.sh are the DLT tape drive device (/dev/rmt/0) and the directory that you want to back up, for example, GROM. backup.sh should be edited to run tar.sh on all of the directories that need to be backed up to tape. Backing up about 10GBs of data requires about 5 hours. The backup.sh script looks like:

tar.sh /dev/rmt/0 ALTO
tar.sh /dev/rmt/0 CAVE
tar.sh /dev/rmt/0 CRAC
. . .

The tar.sh script is in /export/home/field/bin.

19. **Execute backup.sh.** Execute the script. It will look for the end of any data already written on the tape, and then begin writing. Each directory will be written as its own tar file. This is done to prevent a problem with a single tar file from causing all of the data to be unreadable. When each directory is finished the tape will be rewound, fast forwarded to the end of the data on the tape, and the next directory written (this is why the tape **MUST BE** clean as described above in step 17).

20. **Cleanup.** When writing of the tape is finished move all of the written station directories on the workstation disk from /data to some other directory (like /data/ontape) and keep them until the data from the tape has been written to another computer in case the tape becomes unreadable.

21. **Unload the tape and shutdown the workstation.** Press the Unload button on the front of the DLT tape drive. When the green Operate Handle light comes on lift the lever below the tape and remove the cartridge. In an Xterm window on the workstation "su" to root, and enter "sync;sync;halt" (without the quotes, of course). The workstation will shutdown. When the OK prompt appears press and hold the power button until the power goes off. Turn off the monitor, the external disk drive by pressing the illuminated blue button, and the DLT tape drive.

Notes:

-- If you need to know which station's data has been placed on a tape you can run the script size.sh, but it is easier to just keep good notes. This script should be in /export/home/field/bin on the field computer in Grottaminarda.

-- The script tar.sh, and size.sh use awk to do some formatting of the information returned by the tar command. If the file sizes written by tar are large enough there may be some error messages generated by awk (integer overflow errors). These do not affect the operation of the scripts, they are just an annoyance.

-- Both the tar.sh and size.sh scripts create log files of their activity. Thelog file for tar.sh is tar.out, and the log file for size.sh is size.out. They will be created in the directory where they are executed from.

-- To read the data back off of the tapes written using the scripts/procedures in this document on a UNIX/Linux machine just write a script using more tar commands than there are tar archives on the tape like

```
tar xvf /dev/rmt/0n
tar xvf /dev/rmt/0n
tar xvf /dev/rmt/0n
...
```

This will read each tar archive until there are no more. Extra tar commands, for which there is no data left to read, will simply generate error messages. Be sure to use the tape device for your machine which does not rewind the tape after each tar archive is read. /dev/rmt/0**n** in the example above accomplishes that on the field computer in Grottaminarda.

(END)