Data Backup and Archive Submission

Backup procedures

Data files from the numerous systems should be copied to a USB SSD (1 Tb) drive and then to the project iMac computer on at least a weekly basis, then transferred to the MOSAiC Central Storage (MCS) facility on Polarstern for the permanent archives.

The data backup procedures are somewhat time consuming due to the number and variety of instruments and different acquisition systems. This document describes a step-by-step procedure for collecting all the data onto a single backup SSD archive and reformatting where necessary. At the end of each cruise leg, the SSD drive can be brought back to Boulder for further archive. We have two 1 Tb SSD drives (Data_1 and Data_2) which can be rotated for this purpose.

We are required to submit the raw data to MCS on Polarstern. To simplify submission, data folders on the SSD drive and iMac use the SensorWeb/MCS short name for each instrument. For MCS archive submission data can be copied directly from the SSD or iMac to the MCS network drive.

These are the backups that should be done at least once a week. Details follow.

- Data folders from DAS100 for 8 instruments: CO_2 Alicat, fast ozone, Licor 7500, Picarro CRDS, NOx, Tekran, RM Young 2D sonic and Metek 3D sonic.

- Daily folders of `DataLog_User` files from the ship Picarro

- Hourly files from the DMS APIMS (only if it is running).

- Control and Data files from the NO_x and fast O_3 instrument laptops.

- GCMS chromatogram sequence folders from the Agilent computer.

- GCMS Daily and Log files from the LabVIEW computer.

- TEI ozone Calibrator and Monitor files from the LabVIEW computer.

- Dew point data from the Michell chilled mirror hygrometer.

- Picarro CRDS and 10m sonic wind data from the Met City 10m tower.

- Daily folders of `DataLog_User` files from the Met City Picarro.

- Ship met and nav data (1 min avg) from the online DShip Extraction tool.

1) NOAA DAS system

From the DAS100 computer in the APIMS rack, copy data up to the previous day for the following 8 instruments. There's a USB extension cable hanging under the monitor for this purpose. Except for the Tekran, hourly files are stored in daily folders named with the two-digit year and three-digit Julian day: `YYJJJ`. Hourly files for the current day are not yet in a daily folder and do not need to be copied over – they are automatically moved at the end of each day to a new daily folder. For the Tekran there is only one file per day and these can be copied up to the prior day.

Here is a list of the source (`DAS100`) and destination (`Data_1` SSD) directories for backing up each instrument:

```
D:\Data\CO2_Alicat\YYJJJ  → Data_1\CO2_Alicat\YYJJJ
D:\Data\FastO3\YYJJJ      → Data_1\ozone_fast_ship_bow\YYJJJ
D:\Data\Licor\YYJJJ       → Data_1\licor_7500ds_ship_bow\YYJJJ
D:\Data\Metek\YYJJJ       → Data_1\sonic_anemometer_3D_ship_bow\YYJJJ
D:\Data\NOx\YYJJJ         → Data_1\nox_ship_bow\YYJJJ
D:\Data\Picarro\YYJJJ     → Data_1\crds_ghg_ship_bow\YYJJJ
D:\Data\RmYoung\YYJJJ     → Data_1\sonic_anemometer_2D_ship_bow\YYJJJ
D:\Data\Tekran\YYJJJ      → Data_1\hg_ship_bow\...
```

2) Raw data from the Picarro analyzer

The `DataLog_User` files from the Picarro are high rate raw data, saved in daily folders in a nested structure on the Picarro analyzer computer: `~\DataLog_User\YYYY\MM\DD`. There's a shortcut to the `DataLog_User` folder on the desktop. Copy all the new `DD` folders through the prior day and rename these to include the year and month: `YYYYMMDD`. Don't copy files from `DataLog_User_Sync`.

```
C:\UserData\DataLog_User\YYYY\MM\DD...→
Data_1\crds_ghg_ship_bow\_DataLog_User\outbox\YYYYMMDD...
```

Once you have the daily folders some reformatting is necessary. This can be done later with the SSD drive mounted to the iMac computer. The raw `DataLog_User` files are not synchronized to the start of each hour, which is inconvenient for our processing code. The files should be parsed out into exact hourly files to be consistent with our other data streams. Copy all the `DataLog_User` files out of their daily folders in `outbox` to the `inbox` folder. Then run the `parseDatalogUserFiles_Ship.py` script to create the hourly parsed output. The python script is in the `inbox` folder.

On the iMac the lazy way to do this is to open the Terminal app, type `cd` and a space, then drag and drop the `inbox` folder to the terminal window, which copies the full path to the terminal, then hit enter. Then, type `python` and a space in Terminal and drag/drop the python script to the window and hit enter. This executes the script on the current directory, and you should see a group of hourly parsed files appear in the `inbox` directory with names like `ghg0YYMMDD_raw.txt`.

Delete the last parsed file since it's only a partial hour, and copy the rest into the `parsed` folder, and then move them into daily folders. The raw `DataLog_User` files can be deleted from `inbox` when done, except for the last one which has a bit of data for the first hour of the next run and should be kept in `inbox`. For the next run, the first parsed file is discarded since it contains a fragment of data from the last hour of the prior run. Confusing, I know.

You can then compress the daily folders of `DataLog_User` data in outbox to save disk space. For the MCS archive, we will copy the daily folders of Picarro data from DAS100 (step 1 above) and the zipped raw daily folders of `DataLog_User` files. We won't copy the parsed files to MCS since they are contain information identical to the raw `DataLog_User` files

3) DMS APIMS

The DMS mass spec data is saved on the National Instruments PXI controller on the back side of the rack. The SIM data files are downloaded from the controller to the mass spec laptop using the NI Max utility. Open the Max utility, right click on the remote system (NI-PXI8108-2F136D6F) and select `File Transfer` from the drop-down menu.
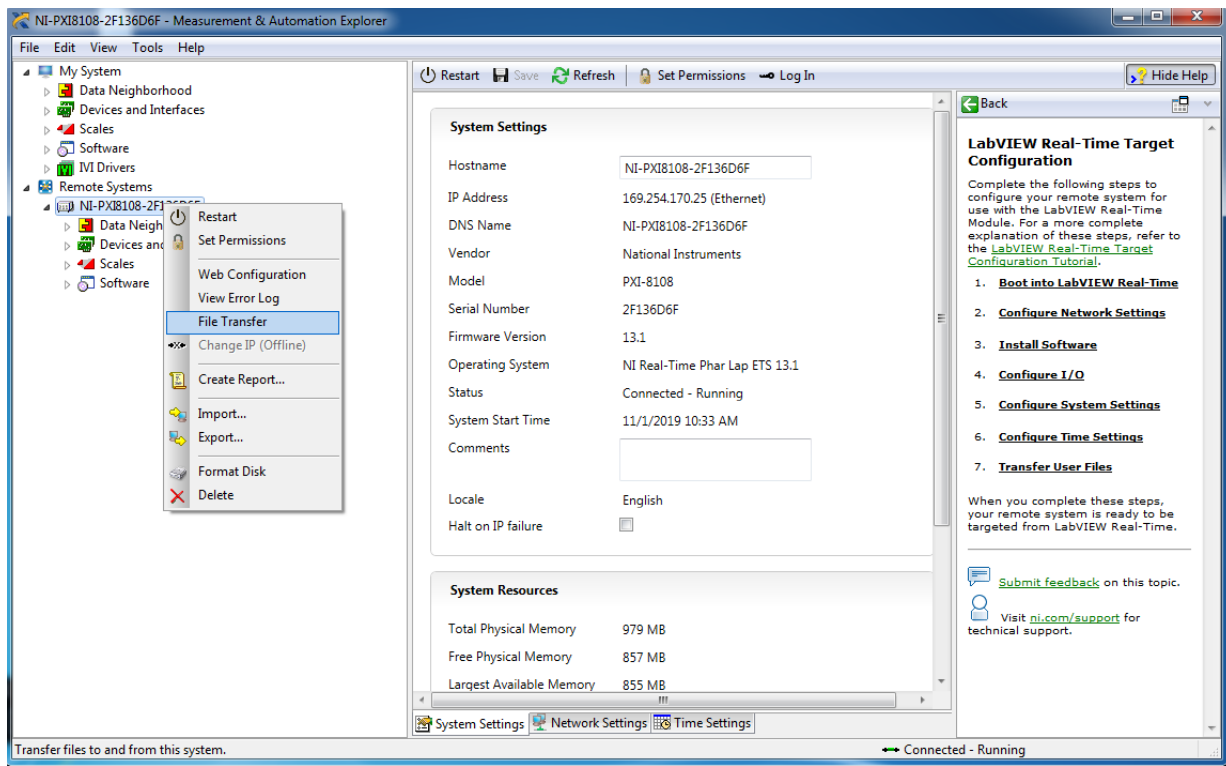
This opens a window showing the remote drive where the `Data` folder resides. DMS flux files are in the `SIM` subfolder. Copy all new hourly SIM files up to the end of the prior day to the laptop directory shown below, and then copy them to the USB SSD backup drive:

```
169.254.170.25\Data\SIM\… → C:\Users\apims\MyDocuments\RTData\SIM\…
                         → Data_1\dms_apims_ship_bow\SIM\…
```

The DMS files are not arranged in `YYJJJ` folders. Create these folders manually on the SSD drive so the data is organized the same way as the DAS100 folders.

4) Michell hygrometer

The Michell logs data to a single continuous file from the time logging was started. The file is on an SD card on the front panel of the instrument. We copy the newest file(s) to a folder on the desktop of the APIMS laptop, and then to the SSD drive. The APIMS laptop is the only computer with a built-in SD card reader (on the left front edge). The file names are a strange format showing the start day, month and time: `DDMMhhmm.txt`. I've been renaming these to something more logical and consistent with what we use for other instruments: `MDPYYYYMMDDhhmm.txt`

The procedure is: first stop data logging on the Michell, then remove the SD card and copy the latest file onto the APIMS laptop and rename it, then copy that to the SSD drive. Don't remove the SD card without first stopping data logging, or the file may be unreadable.

```
S8000\DDMMhhmm.txt  → Desktop\Michell DewPt\ MDPYYYYMMDDhhmm.txt
                    → Data_1\t_rh_ship_bow\ MDPYYYYMMDDhhmm.txt
```

Then check the clock on the Michell and set time if necessary, using the time displayed on the GPS system in the rack just below the Michell. Then insert the SD card and restart logging. The operator's instruction sheet and Michell hygrometer manual have details on how to start/stop logging and set time. But here is a brief description:

There's a large round button in the center of the display, the `menu/enter` button, with arrow keys to either side. Response is a bit slow when you push a button, so wait a second or two each time. To get to the `LOGGING` menu: press `ENTER` key once to get the `MAIN` menu, then down arrow to select `SETUP` submenu, then `ENTER` key again. Then down arrow to `LOGGING` menu, then `ENTER`. Down arrow key to select `LOGGING`, then `ENTER`. Use up / down arrow keys to select `YES` to start logging or `NO` to stop logging, then `ENTER`. Left arrow key to go up to the `SETUP` menu, then down arrow to `EXIT`, then `ENTER`. Now you're back to the normal display. When it's in logging mode a small document icon appears in the upper right corner of

the three output displays on the front panel. The clock is set in a similar way from the `SETUP:CLOCK` menu item.

5) NOx and fast ozone

These laptops are Windows XP and won't mount the 1 Tb SSD, so first transfer files to a USB stick and then copy over to the backup SSD drive on another computer. USB ports are on the rear left side of each laptop and are hard to reach. The NOx system has a USB multiport extension on the back to make connection easier. Here are the source and destination directories for the backup:

```
C:\Documents and Settings\arl\My Documents\LabVIEW DATA\MOSAIC\Fast
NOX#1_data_files\… → Data_1\nox_ship_bow\Fast NOx#1_data_files\…

C:\Documents and Settings\arl\My Documents\LabVIEW DATA\MOSAIC\Fast
NOX#1_control_files\… → Data_1\nox_ship_bow\Fast NOx#1_control_files\…
```

And on the fast ozone laptop:

```
C:\Documents and Settings\arl\My Documents\LabVIEW DATA\MOSAIC\Fast
O3#2_data_files\… → Data_1\nox_ship_bow\Fast O3#2_data_files\…

C:\Documents and Settings\arl\My Documents\LabVIEW DATA\MOSAIC\Fast
O3#2_data_files\… → Data_1\nox_ship_bow\Fast O3#2_control_files\…
```

6) GCMS

Data for the GCMS system is on two computers, the LabVIEW PC running the sample control software (left side of rack behind the pull out keyboard/monitor) and the Agilent computer controlling the mass spec acquisition and control system (right side of rack below the GC oven). The Agilent keyboard/monitor is on top of the left side of the rack. From the Agilent computer, copy over sample folders and LOG files from the MOSAiC data directory up to the current completed sequence run.

```
D:\MassHunter\GCMS\1\data\MOSAIC_GCMS_data\… →
                                Data_1\gcms_ship_bow\Agilent\…
```

From the LabVIEW control computer, copy over daily and log files from the BUP directory:

```
C:\BUP\daily\… → Data_1\gcms_ship_bow\Control\daily\…

C:\BUP\log\… → Data_1\gcms_ship_bow\Control\log\…
```

7) Slow ozone

These files are on the LabVIEW PC that runs the GCMS sampling system, also in the BUP folder. Copy Calibrator and Monitor files up to the prior day.

```
C:\BUP\MOSAIC_ship_O3\49C_Calibrator\… →
                        Data_1\ozone_slow_ship_bow\49C_Calibrator\…

C:\BUP\MOSAIC_ship_O3\49C_Monitor\… →
                        Data_1\ozone_slow_ship_bow\49C_Monitor\…
```

8) Met City Picarro and Sonic Anemometer

Met City wind and Picarro data are logged on Matt Shupe's DAS at the base of the tower. Matt's group transfers the files to their Workstation on the ship every couple days or so. Go to Matt's container on the foredeck to copy them from the Workstation computer. Ask someone from Matt's group to show you how if uncertain. Look in the `D:(Data)` disk. Our files are in the `Meter 10m` and `Picarro` folders.

Unfortunately, their DAS gives all these files the same basic name, beginning with the prefix `msc0,` and they aren't saved in daily folders. Copy the most recent files up to the prior day for each instrument into the root `sonic_anemometer_3D_ice_station_tower_10m` and `crds_ghg_ice_station_tower` directories on the SSD drive. Later, rename them and move into daily folders as described below.

```
D:\Data\Archive\tower\Metek_10m\… →
            Data_1\ sonic_anemometer_3D_ice_station_tower_10m\…

D:\Data\Archive\tower\Picarro\… →
            Data_1\ crds_ghg_ice_station_tower\…
```

Use the iMac computer to rename these files to something more useful. There are two `Automator` shortcuts on the iMac desktop: `msc0_to_son2.app` and `msc0_to_pic1.app`. Drag and drop all the `msc0` sonic files from the SSD drive onto the `msc0_to_son2.app` icon and they will all be renamed with the `son2` prefix. Drag/drop the Picarro files on `msc0_to_pic1.app`, to rename them with the `pic1` prefix. Then organize the `son2` and `pic1` files into daily `YYJJJ` folders on the SSD drive.

The raw `DataLog_User` files from the Met City Picarro need to be transferred directly from the analyzer and reformatted to hourly files as described above for the ship Picarro. Carry a big USB stick out to Met City to get the files since they are large. Again, for the MCS we will only

archive the daily folders from Matt's DAS and zipped daily folders of the `DataLog_User` files, not the parsed hourly files.

9) Ship Met and Navigation Data

Ship data is available on the network from the main webpage at http://www.fs-polarstern.de/index.php?lang=EN&aktion=main.

Select the `DSHIP` tab, and then `DSHIP Extraction` at the bottom of the page. Click `Data Extraction` to make a request. It's possible to set up a template for the variables you want to download on a regular basis. I've done this on the iMac. Click `Load User Template`. My user name should come up, if not type in `bblomquist` for user. Then select the `ShipData2` template. If you have any issues with this you can also set up a template from scratch with your own user name and password, which the AWI data folks show you how to do. The DSHIP template should select the following variables at 1 minute time resolution:

```
Magnetics:gyro heading
Magnetics:gyro pitch
Magnetics:gyro roll
Trimble 2 GPS:course made good
Trimble 2 GPS:position latitude
Trimble 2 GPS:position longitude
Trimble 2 GPS:speed made good
Weatherstation:air pressure
Weatherstation:air temperature
Weatherstation:ceiling
Weatherstation:dewpoint
Weatherstation:direct radiation
Weatherstation:global radiation
Weatherstation:relative humidity
Weatherstation:relative wind direction
Weatherstation:relative wind velocity
Weatherstation:true wind direction
Weatherstation:true wind velocity
Weatherstation:visibility
Weatherstation:water temperature
```

When variables are selected, fill in start and stop times at the top of the variables table, be sure 1 min interval is selected, and click the `Next` button. I usually request whole days of data, from 0000 UTC to 0000 UTC. Click `Next` once again on the `Parameter Settings` tab. On the `Additional Format Settings` tab enter user defined values of 'NaN' for error types at the bottom of the page, then click `Next` again. On the `Order Settings` tab, enter a

filename like `2019-11-14_to_11-20_ShipNavMet` indicating the date range for this extract and submit the request. It usually takes 10 min or so, then you can go to the `Extraction Download` link on the `DSHIP Extraction` page to get the file. It comes with a `.dat` file extension, which I change to `.txt`.

The DShip download file is not saved in UTF-8 format, so python can't read it. Double click the file to open it with the BBEdit editor. The bottom of the editor window has a drop down menu to change the encoding to UTF-8. Then save.

The last step is to parse the single download file into daily files that are easier for us to use. Copy the downloaded file into the `/Users/fluxuser/Documents/MOSAIC_2019/ship_data/inbox` directory on the iMac and run the python script `parse_Dship.py`. The procedure is the same as for parsing the Picarro `DataLog_User` files described above. You should see a group of daily parsed files appear in the `inbox` directory with names like `dsp0YYMMDD_raw.txt`. Copy these files into the `parsed` folder, and copy the DShip download file into `outbox`. Matlab scripts will read the parsed files to get the ship data. We don't archive any of these files to MCS.

10) Copying files from the SSD to the iMac

To synchronize files between the backup drive and the iMac use the `DropSync3` app to run an `rsync` session. Attach the SSD drive to a USB port and open the `DropSync3` app – 7th icon down on the left hand sidebar. This opens a window showing the iMac `MOSAIC_2019` directory on the left and the SSD `MOSAIC_2019` directory on the right. The two directory structures should be identical, but in this case the SSD side will contain new or updated files. To sync the new files over to the computer, click the arrow icon pointing from the SSD side to the computer side. It first runs a comparison showing you the new or updated files, which you can then execute to copy files to the computer directory from the SSD.

This procedure is really fast compared to copying manually. You can also sync back from the computer to the SSD drive to sync updated Matlab scripts or analysis plots and data from the `_matlab` or `_scientific_analysis` folders.

DropSync copies over new files and directories and over-writes old versions with newer, but will not delete files from the destination side that don't exist on the source side. To delete files permanently, remove them manually from both directories before running DropSync. If you delete from one side only, the files will be copied back the next time you do a reverse sync.

11) Submitting to MCS Archive

The ship web page has instructions for mounting the network MCS drive on your computer, or the iMac has the smb://mcs.fs-polarstern.de/platforms link in the `Go:Connect to Server` menu.

Our ship-based instruments are all in the MCS directory,

`platforms/vessel/polarstern/bow_tower/`

with subfolders for the following instruments. In each case, raw data goes into a subfolder in the `exdata` subfolder. Note the `CO2_Alicat` is not in MCS since this is just QC data for the Picarro.

`crds_ghg_ship_bow, dms_apims_ship_bow, gcms_ship_bow, hg_ship_bow, licor_7500ds_ship_bow, nox_ship_bow, ozone_fast_ship_bow, ozone_slow_ship_bow, sonic_anemometer_2d_ship_bow, sonic_anemometer_3d_ship_bow, t_rh_ship_bow`

The MCS directory for the Met City Picarro is

`/platforms/pack_ice/ice_ps/flux_tower_12m_ucb/`

`crds_ghg_ice_station_tower`

Otherwise, the subdirectory structure and data files for this Picarro are the same as for `crds_ghg_ship_bow`.

The MCS directories for the flux chamber systems are:

`/platforms/pack_ice/ice_ps/dms_chamber_portable`
`/platforms/pack_ice/ice_ps/co2_ch4_chamber_portable`
`/platforms/pack_ice/ice_ps/ozone_chamber_portable`

Within the `exdata` subfolders you will find further 'science activity' subfolders with names like `PS122-1_1-70` for the actual data, where the '`PS122-1`' is the cruise and leg number, '`_1`' indicates continuous measurement, and I don't know what the '`1-70`' means. For continuous measurements like most of ours, one science activity folder will be created at the start if each leg when you enter a `DShip_Action_Log` 'station start' event in the system, indicating the beginning of a new continuous measurement period. At the end of a leg a 'station end' event is entered for each continuous device and a new 'station start' is immediately entered to begin the new leg.

Note that for discontinuous measurements, like flux chamber deployments, there's a new science activity folder name created each week, so the exact subfolder you use for a given measurement depends on the week it was done, and these folders are created automatically when you register a sample collection event using the Floe Navi interface, or register the sample after-the-fact on DShip. Clear? No? Ask AWI data support when you have questions.

When you've found the correct subdirectory on MCS, just copy/paste the files over (cmd-c and cmd-v on the Mac). Take a look at the files already in the archive and you'll see how it should go. Note, the upper level `bow_tower` and `flux_tower_12m_ucb` folders also have an `exdata` subfolder, but these are not used since the tower itself produces no data.