

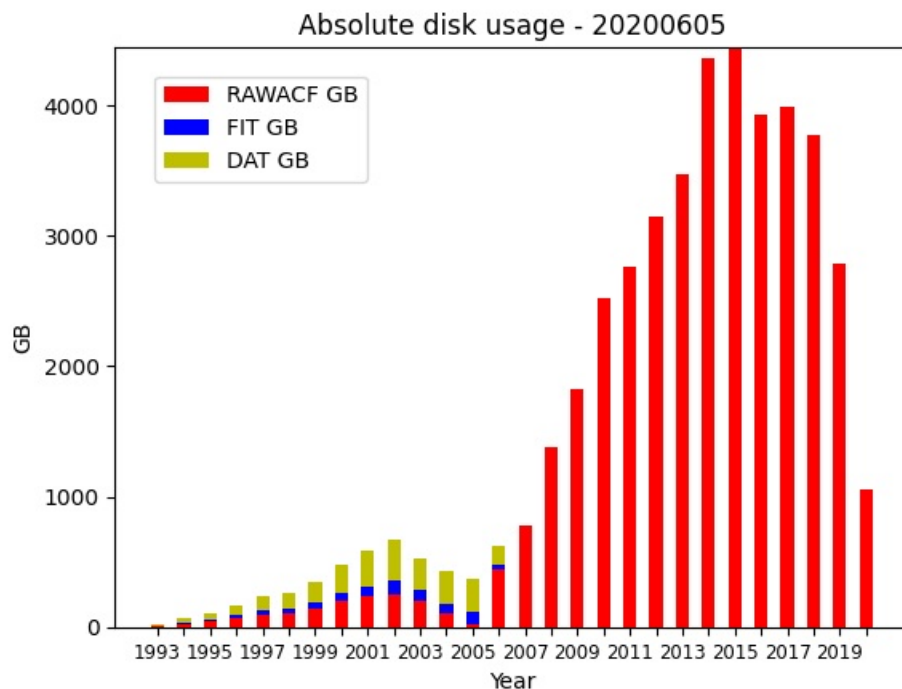
# DDWG Report: SuperDARN Workshop 2020

## Activity over the past year

1. Still doing online distribution, no hard drives requested
2. VT has purchased and set up two new NAS systems in Jan/Feb 2020, each has roughly 73TB capacity. The group expects they may be able to purchase an additional two NAS systems to contain 1993 to 2018 data, this data is currently on two RAID arrays set up in 2012/13, with at least one failed disk among them.
3. Sessai/Yuka retrieved DAT files
4. Judy retrieved DAT files
5. 'Deletions' directory made available on Globus
6. hdw.dat files responsibility passed to another working group
7. Data management audit performed, October 2019, please see a summary [here](#)
8. Downtime summary:
  1. BAS mirror down temporarily from June 5th to 11th
  2. VT had downtime affecting data access for 2019 data, on October 31, 2019 to add more storage
  3. USASK temporary mirror had a hard drive failure & network switch failure in early March 2020
  4. USASK Globus mirror had downtime from March 16 - 22 2020 for upgrades
  5. USASK Globus mirror had downtime from March 23 - 30 due to filesystem issues
  6. BAS mirror had a downtime on May 8 2020 from 08:00 - 13:00 UTC for maintenance
9. Gap resolution done in October 2019 for September 2017 to August 2019 data inclusive
10. Issues brought up on mailing list (\* - not yet resolved):
  1. \*RAWACF files produced by `dat2orawacf` with low-power thresholding issue
  2. \*Nonsequential DAT files (records are out-of-order in time)
  3. Two blacklisted Canadian files from RKN/SAS due to main computer failure at RKN and old MAIN computer from SAS shipped to RKN
  4. \*Still 2255 DAT files to be recovered (most from 2000-2006, 14 files from 1995-1998)
  5. DAT hashes file on Globus issue
  6. A BPK file with a timing error was blacklisted
  7. A RKN file was blacklisted, as it was corrupt on the QNX filesystem, and not recoverable.
  8. Clyde River Borealis files were removed from the distribution, as they did not follow a naming
  9. DCE files were blacklisted from July 5 and July 6 2018 due to failures in the Phasing Matrix. This is in addition to files from June 22nd 2018 - January 28th 2019.
  10. A power outage at Kapuskasing caused some files to be corrupted from 20191208.17 to 20191210.08, they were blacklisted.
  11. One JME file from 20200102.0000 was updated standard. They are currently in a separate directory (along with Saskatoon and Prince George data)
11. DDWG Mailing list updated with a representative from every group (*Still waiting to hear from Australian group*)
12. Future work: See the 'Action items' below

## Data summary

### Absolute disk usage by year and data type



The above plot shows how many GB of storage are in the main distribution directories for RAWACF, FIT and DAT files. It is calculated using the command-line program `du -s`. The title of the plot shows the year, month and day the plot was generated. Note that this includes RAWACF files that were generated from DAT files.

## Meeting summary, June 3 2020 12:00 - 15:25 UTC

Please see the minutes/agenda for the meeting [here](#)

Please see relevant files on the DDWG github page [here](#)

### Action items

1. Kevin Krieger/USASK:
  1. Share results of audit (please see summary [here](#))
  2. Reconcile issue numbers on github with historical issues
  3. Check consistency of responsibilities section with the PI agreement
  4. Check with DSWG about data level definitions, to make sure we are consistent with their docs
  5. Keep track of which files were converted from DAT to RAWACF and place in separate directory to 'dat'
  6. Make available a list of 'fixed' files alongside blacklist and failed files lists
  7. Add step 6 for gap resolution procedure: Inform DDWG of old files added
  8. Add to charter old file formats such as FIT that are also in the distribution (if any)
  9. Send message asking group to look for missing DAT files and original DAT files from the issue Evan outlined with low-pwr thresholding
2. DAWG:
  1. Provide tool for checking DAT files similar to the tool for checking RAWACF files
  2. Provide tool for checking RAWACF files that were converted to DAT with low-pwr issue
  3. Provide tools for fixing blacklisted files
  4. Provide guidance on possible fixes to DCE files that had phasing matrix issue
3. Kevin Sterne/Paul Breen/DDWG members who stage files for them:
  1. Work to get all PI institutions' radar data flowing to both VT and BAS
4. Data Policy Task Force:
  1. Who is responsible for fixing files?
  2. Should the DDWG be placing higher level data products in the scope?
5. PIs:
  1. Inform DDWG of appropriate responsibilities and policies
  2. Provide access to an updated PI agreement
  3. Discuss the current roles within the DDWG, should they be evolving?:
    1. Role of VT - Kevin Sterne handles getting data from many PI groups, automated except for tracking down data issues
    2. Role of BAS - Paul Breen handles getting data from many PI groups and maintaining mirror, this is a risk, as we have heard from BAS that they don't have any resources for development
    3. Role of USASK - Kevin Krieger maintains the Globus mirror and has been maintaining a *temporary* repo for ~3 years so that other groups would have time to transition to Globus. This temporary repo is a risk (failing hard drives, manual tasks) and an

inefficient use of resources and I recommend it to be decommissioned ASAP

4. Discuss potential future servers within the SuperDARN Community. Dartmouth has expressed interest

6. DSWG:

1. Add categories for blacklisting to level 1 standard documentation

2. Provide guidance on data product level definitions (level 0, level 1, etc)

#### **Other topics discussed**

1. Charter/policies/procedures draft

2. Ease-of-use of Globus vs BAS repositories, some users have trouble with Globus, Kevin Krieger willing to help

3. Why are the USASK and BAS repositories not synchronized? Issues at BAS with retrieving data from USASK, issues at USASK with temporary mirror

4. Where should file checks be done? Currently done before files go onto repositories, but could and should be done on the user side as well.

5. Discussed possible SuperDARN-wide funding to have a dedicated resource for data distribution