

Data Distribution Working Group

June 4th 2018

Kevin Krieger¹, Kevin Sterne², Paul Breen³

SuperDARN Workshop
Banyuls Sur Mer, France



1 – University of Saskatchewan

2 – Virginia Tech

3 – British Antarctic Survey



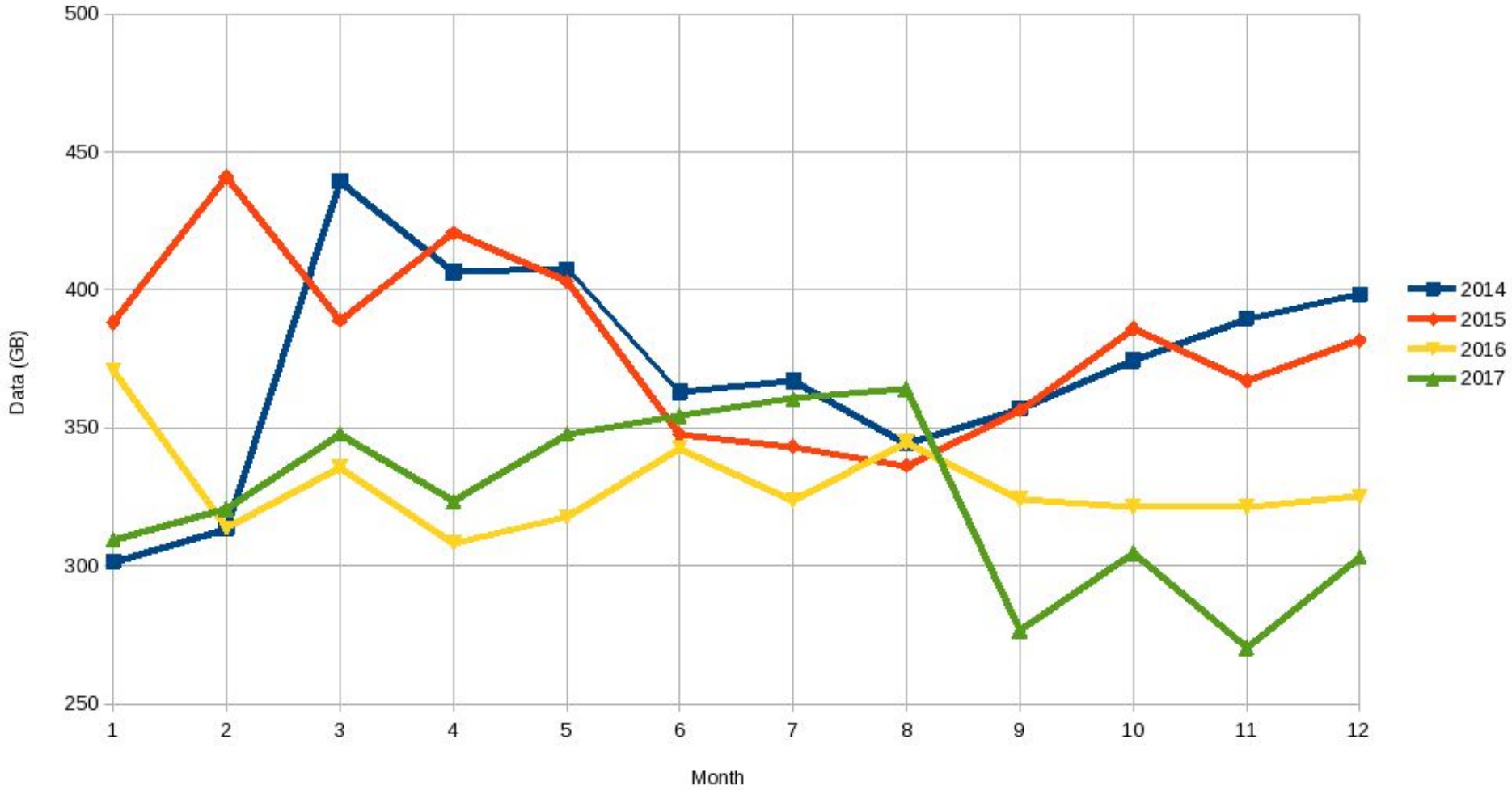
DDWG: Summary

- . Full online data distribution in 2017-18
 - No hard drives written or distributed
- . Expect possible interruption in service on VT servers
 - more hard drive space required within next year
- . Globus used for data transfer
 - Attended Globusworld conference April 25-26
- . Continuing support for temporary (latest months only) data server `superdarn-cssdp.usask.ca` but you should really move to **Globus**

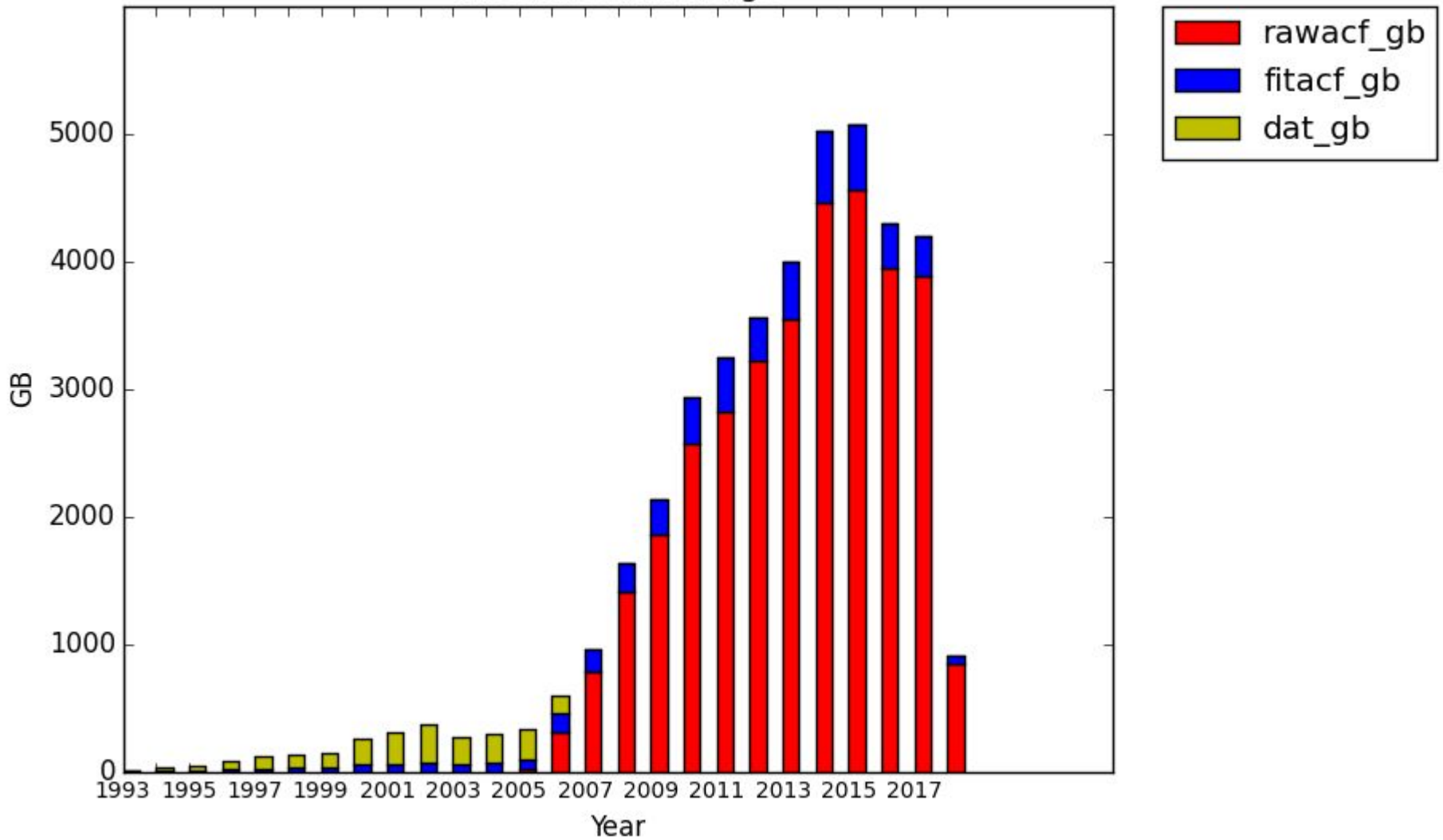


Monthly rawacf.bz2 data available on Globus

UofS SuperDARN Data Mirror



Absolute disk usage



University of Saskatchewan Data Mirror

- Temporary mirror still accessible
- Globus mirror with full dataset available – contact **Kevin Krieger** – **kevin.krieger at usask dot ca**
- Data from 1993 – April 2018 available
- Mirror downtimes over the past year:
 - April 26th 13:00 – 17:30 UTC 2017 – moved temporary server at University of Saskatchewan
 - December 3rd – December 4th 2017 – Server side credentials expired



Issues & Discussions

- List of blacklisted files is available at:
 - `~/sddata/.config/blacklist/` on data mirror
- Known data problems listed on DDWG website
 - blacklist all files with known problems.
- Non sequential DAT file issue
 - timestamps not sequential or duplicated in two different files
 - not addressed yet



Issues & Discussions 2

Issue #	Date	Description
12	June 27th 2017	LYR files prior to public release
13	June 28th 2017	Blacklisted hok files with no data 201701
14 through 22	Mid July 2017	Blacklisting files with known issues on DDWG page
23	July 18th 2017	Corrupted Stokkseyri file
24	July 20th 2017	Corrupted PGR file
25	August 1st 2017	SAN files with no minute in filename
26	Nov 7th 2017	526 Corrupt rawacf files - see backscatter package: https://github.com/SuperDARNCanada/backscatter

Globusworld 2018

April 25th - 26th 2018, Chicago, IL

- 1/3 of R1 Universities use Globus now
- **Make sure we acknowledge Globus in journal articles/publications/websites**
- Future roadmap for Globus developers
 - trigger-action programming (automated secondary product generation?)
 - Improving performance
 - automation of publications/groups/web apps/etc...



More Globus information

- <https://www.globus.org/>
- Script utilizing python SDK for SuperDARN data:
 - <https://github.com/SuperDARNCanada/globus>
- New Command line interface: <https://docs.globus.org/cli/>
- Python SDK: <http://globus-sdk-python.readthedocs.io/en/stable/>



Future

- Globus storage currently moving from *remora5.sharcnet.ca* to *graham.computecanada.ca*
 - Holding up data uploads to Globus mirror
- “Standard” set of Canadian FITACF files to be on Globus
- <https://www.frdr.ca/repo/> for a potential place to publish ‘finished’ datasets
 - archiving for publications; version control; assigns DOI number to dataset to refer to in publication
- Globus web-applications
- For more up to date information on the DDWG see:
 - <http://vt.superdarn.org/tiki-index.php?page=Data+Dist+Working+Group>