DDWG Meeting Minutes

May 26 2021 12:00 UTC - 13:45 UTC Virtual Webex

Attendees

Name	Institution
Kevin Krieger [Organizer/Minutes]	USASK
Marina Schmidt	USASK
Kevin Sterne	VT
Simon Shepherd	Dartmouth
Evan Thomas	Dartmouth
Jordan Wiker	JHU/APL
Enrico Simeoli	IAPS/INAF
Paul Breen	BAS
Tim Barnes	BAS
Jianjun Liu	PRIC
Judy Stephenson	UKZN
Tomoaki Hori	ISEE
Nozomu Nishitani	ISEE
Aurélie Marchaudon	IRAP
Tim Yeoman	Leicester
Mikko Syrjäsuo	UNIS
Mike Ruohoniemi	VT
Kathryn McWilliams	USASK
Federica Marcucci	IAPS/INAF
Pasha Ponomarenko	USASK

Brief update

A brief update was given for how much data is in the main distribution

How to access data with Globus (Kevin Krieger) or BAS (Paul Breen)

Major updates from the past year were shown:

- SimonS: Question about list of failed files and list of updated files, are they on the BAS mirror as well as the Globus mirror?

- KevinK: These lists are specific to the Globus mirror

- PaulB: BZ2 corruption issues for example are available and captured on the BAS mirror and available through an API call

- NozomuN: ISEE University working on copying files - starting from 2012

and will soon complete this and hopefully will also start service with a mirror in the near future.

Discussion

File delay discussion

- SimonS: Question about how files go from USASK to BAS. There seems to be a 2 week daily for all CDN radars
- KevinK: We stage them up daily for BAS. The 2 week delay is likely a one-time problem, I will need to look into it when I have time (likely next week).
- KevinS: What is an appropriate timeline for a delay? Some radars have issues with getting data back daily
- EvanT: Can get IMF data daily
- SimonS: What is your target(for delay to mirrors)? Discussion about 2 hours?
- KevinK: 2 hours target not related to map files or everyone in the community, it's an ideal to strive for but not possible for everyone.
- KevinS: Nothing in the PI agreement about a target.
- KevinK: We need to come up with a target that is appropriate for everyone in the network.
- EvanT: Discussion about map files is a bit beyond our scope right now - everyone doing best-effort transfers
- SimonS: Shouldn't be a problem for everyone to get data uploaded on a daily basis
- KevinK: Problem with CDN files is likely due to power outages, server failure, etc over past month, bear with me as I find time to look into it.
- PaulB: Similar to Kevin K. our scripts run daily, there's always exceptions to the daily rule though and there is about a 5 day lag for things to work their way through the pipeline
- KevinK: Update for RKN - computer failure and trouble getting a hold of site custodian.

Automated emails discussion

- SimonS: Would be helpful to have more communication on WG if there are issues. Would be appreciated.
- KevinK: Everyone feel free to email list whenever there is an issue.
- KevinS: Found it helpful when Paul would reach out when there were data issues/delays. I realize this creates extra work that may be hard to manage. Can this be taken on and see how it goes?
- PaulB: Did this more often years back, it just comes down to time now. This is done by automated alerts (Nagios). Cryptic messages could be

made more human-readable and sent to individuals who are responsible for those radars' data. Need to find time to do this. Can we have a recommendation for how long the delay threshold is before sending an alert?

- SimonS: Wouldn't mind receiving an email daily, comes back to what data mirror operators can manage. Even cryptic email would be helpful
- PaulB: If you respond to email - it will acknowledge the alert and clear it.
- MikkoS: Network of radars so there should be a bit of responsibility placed on data managers to help out the server operators. Willing to work on this.
- KevinS: Going back to threshold, is that something the PIs can come to concensus on?
- SimonS: two thresholds: one to trigger alerts (1 day would be good, but may be too fast), and one for lag/delay from data on local server to mirrors (1 day reasonable except for some radars)
- KevinS: Second delay, can you clarify?
- SimonS: Sync daily with rsync to BAS, so check delay daily. 2 week delay with CDN radars, why not 1 day?
- KevinS: 2 week delay sounds like a temp problem
- KevinK: Yes it's a temp problem.
- MarinaS: 3 power outages in last two months
- KevinS: So that's a temporary issue. Goose Bay data after email sent took a day or so to get through system. couple day delay
- EvanT: Seems like delay from radar site to institution seems unavoidable, seems like Simon is talking about delay from institution to mirrors.
- SimonS: Once reminded gbr was quick to be uploaded, KevinK seemed surprised, need someone to be checking on these things.
- KevinS: Needs knowledge on KevinK or PaulB side about radars otherwise they can't know. It's not automatable.
- EvanT: Goes back to Mikko's point, it's also on the data managers.
- MarinaS: globus mirror is at the same point that BAS is.
- KevinK: Again, this is a temp problem.
- SimonS: Comes back to data managers to communicate better.
- KevinK: How much work to get automated emails going Paul?
- PaulB: Just need to change contact in Nagios (ideally). Can do this. Issue we have is that some of the alarms say "Check logs for details". Don't expose these logs but may be able to workaround this. At least this shows you there is a problem.
- TimB: Changing to user friendly comms may be doable. Can look into in future.
- MikkoS: Would like to only see stuff related to our radar, not everyone's.
- PaulB: Each check related to specific radar so this is possible.
- MikeR: Since scripts are radar specific, can you customize delay for each radar?
- PaulB: Yes that is possible. Some radars have higher inherent lag than others.

- MikeR: Daily notification for us may be too frequent, perhaps weekly. For others it may be useful for daily.
- PaulB: Once threshold is passed - alert is sent, then again an alert will be sent after the threshold is passed again, unless you reply to silence the alert.
- MikkoS: may be possible to change this in Nagios, depending upon how you're using it.

Gaps resolution discussion

- KevinK: Do people still find this useful? I receive maybe half responses, should I still do this on a yearly basis or even more periodic?
- AurelieM: Very useful
- KevinS: This is useful
- PaulB: Very useful - forced to look at this. Development of temporal coverage app is also useful to see visually the gaps. This could be a more proactive way for data managers
- KevinK: Maybe along with the emails I send can also send them a link to the visual temporal coverage app. Discussion at usask to do the gap resolution more than once a year to help with finished data sets being uploaded to FRDR - Federated Research Data Repository.

Data managers uploading data to VT and BAS

- KevinK: We agreed last meeting to upload all radars to both VT and BAS for redundancy.
- SimonS: Not sure everyone is doing that. Explained in several emails that I shut off the transfer to VT temporarily until we sort out other issues.
- KevinK: Is there a reason to shut off the transfer?
- SimonS: It's a hassle to figure out where the data is in order to make map potential files, trying to reorganize things in anticipation of iceland build. Christmas valley data more up to date than quite a few radars in system
- KevinS: Also guilty of that, still todo to get Paul login.
- PaulB: No problem, I had this on todo list as well.
- SimonS: You will also lose kod, mcm, sps, potentially ksr.
- KevinK: Who will lose access?
- SimonS: VT will lose access when Bill needs to rearrange system. He is automatically uploading to BAS right now.
- KevinS: Will work with Bill to keep transfers going.

Data gaps discussion

- KevinK: Can speak to CDN data files -> still uploading one month at a time Borealis data. Australia radars Jim let me know they had issues. I

wasn't aware that adak had been decommissioned.

- SimonS: Decommissioned more than a year ago
- KevinK: Another data point that shows we need better communication.
- AurelieM: What about PI reports every year? Can the info be sent to WGs?
- KathrynM: Not clear to me that information from PI reports to be shared with WGs. Can check with PIs at meeting.
- AurelieM: Can have template for the reports? Radar on? etc?
- EvanT: Radar audit, can this be sent out internally within SD? Has general info about status of radars, could be useful.
- KevinK: I would find it helpful
- PaulB: Me too.
- KathrynM: You have that as an action item for PIs?
- EvanT: yes I do. Would that address your point Aurelie?
- AurelieM: Yes, maybe too much info than needed by DDWG, but some info could be extracted for the WG.
- EvanT: Could hide certain things in excel spreadsheet.
- PaulB: From my perspective would be helpful to know when radars fully decommissioned, can just turn scripts off for it. For extended downtimes can do something else.
- MikkoS: We do have PIs that can tell us this information, we don't need a formal report, just an email would be useful. Let's leverage the network.
- KathrynM: Tough to know how to rely on PI or datamanager when things change - example of Nagatsuma-san being out as PI now.

Delay threshold discussion

- KevinS: Going back to the delay threshold before an alert is sent, what would be appropriate?
- PaulB: [showing dashboard for BAS superdarn data] - can likely use this dashboard to determine a proper threshold for each radar.

Data quality issues discussion

- EvanT: Haven't touched on data quality control and blocking of files due to any variety of reasons. Something that's always happening is it worthwhile to separate out a working group or task force to work on these issues?
- KevinK: Agreed - I would have thought data standards WG had a better scope for this problem. Always seen it as DDWG simply blocks files if they are told to and if a file is corrupt then just don't put it into the distribution. Saw this as a DSWG task.
- EvanT: Problem is that DSWG is taking top-level approach, and the WG is inactive and they never reached that level of detail. Feels like it's outside

- the DDWG scope, but also not productive to put it under the inactive DSWG.
- MikeR: Something to ask the PIs about, where to locate the responsibility to check files for consistency. Is there a separation now of how to check data files for issues?
- EvanT: There are some checks automated before being placed on mirror but other issues that are harder to automate are more difficult (example, can you make a FITACF from a RAWACF without issue?)
- MikeR: Also some can process fine but contain garbage.
- EvanT: One example is the dat files that have timestamps out of order.
- SimonS: Correct but it also breaks RST when you try to trim files. Not sure about the file checking, thought that was the primary reason why VT is in the system?
- KevinK: I can clarify, VT is grabbing files due to historical reasons and it's a nice division of tasks between us.
- SimonS: Isn't it automated?
- KevinK: The work is to maintain the automation, follow-up with datamanagers when there are connection issues, etc.
- KevinS: KevinK handles outflow (distribution) and I handle ingress of data. So KevinK has contacts for logins, I handle contacts for getting data.
- SimonS: Isn't this where delays are coming from? Seems complicated.
- KevinS: Doesn't seem complicated here. We do checks for files being too small or bz2 issues. KevinK has been checking records of dmaps but haven't had the time to do that here.
- PaulB: We do the same basic checks that KevinS does. Can't do the extra checks that KevinK does because of time and IT issues. It would be very useful to have governance around this, what types of checks should we be doing?
- KathrynM: This can be discussed at PI meeting and put it in the agenda, This involves several WGs. More difficult now because we have various types of radar systems and need to come to common file format that is high quality.
- MikeR: There are some checks that are easy to do automatically, but others are difficult to do automatically. Is there something we can do across collaboration that we can do this efficiently?

Paul sent link to listing of checks they perform: here

Kevin's note: the checks that USASK performs are described here

Meeting end 7:45AM