

GHRC UWG Report from the Onsite Meeting October 20-21, 2016
National Space Science & Technology Center,
University of Alabama – Huntsville
Huntsville, Alabama

Executive Summary

The GHRC provided a full day of presentations that addressed the recommendations agreed upon in the 2015 GHRC UWG report. The discussions consisted of seven sessions covering the following: 1) Introduction; 2) Potential new data sets; 3) Data publications, HyDRO 2.0 and new system architecture; 4) Virtual collections, data recipes, and web and social media; 5) Field Campaign explorer, Python libraries; 6) Event Explorer (Giovanni); and 7) User characterization.

The UWG continues to be impressed by the progress the GHRC has made on the recommendations provided in the first two annual reviews, specifically towards building a unified theme of the DAAC around storm-induced hazards, fulfilling that vision through current and future data holdings, completing a detailed data life cycle process, and developing tools that will allow users to more easily interact with those holdings through data bundles. Of the 20 recommendations provided by the 2014 and 2015 UWG reports, only eight remain for 2017. Many of the 2015 recommendations have been successfully closed, while others were addressed at least in part. Aside from continued movement toward closing all recommendations made by the UWG, the UWG would like to place special emphasis on the development of a 5-10 year plan for sustained success at the GHRC.

Two new members of the UWG are needed for next year's meeting to replace members whose UWG terms are expiring. Some recommendations are provided for in this report.

Meeting Report

A one and one-half (1.5) day meeting was held to review the progress towards meeting defined goals from the previous year's UWG meeting. Fifteen (15) UWG members attended. The expertise, names and e-mail address of the members are provided for in Table 1. In addition to DAAC staff and UWG members, Drew Kittel and Steve Berrick from NASA's Earth Science Data and Information Systems (ESDIS) Project; Steve Kempner (GES DISC), Bob Downs (SEDAC), Suresh Santhana Vannan from Oak Ridge National Laboratory (ORNL) DAAC, Nettie Labelle-Hamer (ASF DAAC), and Anirudh Prabhu from Rensselaer Polytechnic Institute (RPI) attended the meeting.

As in previous years, the first day consisted of several presentations discussing the effort made by the GHRC to address the recommendations provided by the UWG in their 2015 report. The second day consisted of a closed-door session (attended only by UWG members and the GHRC manager) during which the UWG evaluated GHRC progress on the 2015 recommendations, and decided what, if any, new recommendations were to be made in this report.

Expertise	Name	Email
Applications	Kel Markert	km0033@uah.edu
	Robert Griffin	robert.griffin@nsstc.uah.edu
Global Precipitation Measurement	David Wolff*	david.b.wolff@nasa.gov
	Walt Petersen	walt.petersen@nasa.gov
Hurricanes	Jonathan Zawislak**	jzawisla@fiu.edu
	Haiyan Jiang	hajian@fiu.edu
	Stephanie Stevenson	sstevenson@albany.edu
Lightning	Dennis Buechler	Dennis.E.Buechler@nasa.gov
	Eric Bruning	eric.bruning@ttu.edu
	Michael Peterson	michaeljp24@gmail.com
	Steve Goodman	steven.j.goodman@noaa.gov
Passive Microwave	David Duncan	dduncan@atmos.colostate.edu
	Christian Kummerow***	kummerow@atmos.colostate.edu
Severe Weather	Andrew Molthan	andrew.molthan@nasa.gov
	Chuntao Liu	cliu5@tamucc.edu

Table 1: List of 2016 UWG members, their area of expertise and email information: David Wolff* (NASA) was the Chair; Jonathan Zawislak** was the Co-Chair; and Christian Kummerow*** was the outgoing Chair from 2015. Note that names highlighted in red represent members who will cycle out of the UWG after this report has been submitted. Those names highlighted in green are new members of the UWG.

Disposition of Previous Recommendations

Table 2 provides a list of the 2015 UWG recommendations as well as their disposition following the 2016 UWG meeting. Each of these recommendations were either closed, open (with a new recommendation number) or merged into a new recommendation. Of the 20 recommendations from 2015, a total of eight new recommendations remain, showing the significant progress the GHRC management and staff have made since the 2015 UWG meeting.

Recommendation	Description	Disposition
1	Evaluate and update the GHRC mission and objectives in coordination with the UWG members, NASA ESDIS, and Program Managers at NASA HQ.	Closed
2	Develop a 5-10 year vision for GHRC and ensure the new website reflects that message.	Closed
3	GHRC should hold AMS and AGU town halls, develop and distribute information brochures that describe their capabilities to potential data providers (e.g. field campaign PIs) and data users, utilize the NASA hyperwall, and pursue other opportunities (BAMS) to enhance GHRC visibility once the 5-10 year vision is developed and the web page reflects these objectives.	Open. New Recommendation #1
4	Carry out dataset holdings analysis and create a reporting structure that categorizes what is available at GHRC and possibly elsewhere. This compilation should enable prioritization of efforts that will fill the most significant data voids, where these efforts align with the new GHRC mission.	Merged to new Recommendation #2
5	Update public dataset information pages to include data holding analysis results that might be helpful to the user community	Merged to new Recommendation #2
6	Determine a set of useful user metrics, with feedback obtained from the UWG that can be routinely updated and made available to the NASA sponsor, UWG and broader community. Analysis of these metrics should inform the 5-10 year plan	Merged to new Recommendation #2
7	Review the "NOAA Procedure for Scientific Records Appraisal and Archive Approval" (https://www.ngdc.noaa.gov/wiki/images/0/0b/NOAA_Procedure_document_final.pdf) and the PODAAC Data lifecycle (http://podaac.jpl.nasa.gov/PO.DAAC_DataManagementPractices). Assess whether these procedures or a modified version of them are useful formalizations that would aid in creating a data lifecycle plan for existing and future GHRC data holdings.	Closed
8	Create a data lifecycle process for GHRC that can be applied to current and future holdings. Ask NSIDC and PODAAC for their policies and assess utility within GHRC. Publish the data lifecycle on the website, along with a contact, to provide clarity on the process for investigators interested in providing data.	Open. New Recommendation #3
9	Assess what might be useful in the NODC netCDF data template and develop some guidelines or work flows for GHRC to handle future field campaign data.	Closed
10	Develop a data maturity model for GHRC data. Provide this on website and include maturity information for each dataset provided. Review NOAAs data maturity model (http://www1.ncdc.noaa.gov/pub/data/sds/maturity-table-6level.pdf) as a starting point	Closed
11	Recommendation #: Determine LIS technical specifications for data products, latency, formats, etc. Publicize this future data source at appropriate venues.	Open. New Recommendation #4
12	Develop a single tool that can provide broad use to multiple field campaigns and data types.	Open. New Recommendation #5
13	Update the cite our data webpage to include DOI in all the examples given and include a link to the cite our data page on individual dataset information pages	Closed
14	Communicate with the LPDAAC to understand their transition to HTTPS process. Provide highly visible examples, links to examples via email, and as much visibility as possible to ease the transition. A page with examples of different methods to download data, accompanied by example code, would be helpful.	Closed
15	Look at netCDF4 as an internal data format, define common CF-compliant metadata for each data type, and develop tools that will check for metadata compliance.	Closed
16	Explore and identify future users of possible mobile apps for NRT data. An assessment of how GHRC ingests format requirements could be used to broaden app utility.	Closed
17	Create data bundles for scientists who want to study processes. Demonstrate such bundling capabilities for review by the UWG.	Open. New Recommendation #6
18	Develop an attractive visualization that goes along with the new mission and vision statement that would help audiences associate the GHRC with its vision and mission statement.	Closed
19	Discuss the possibility of getting land data from the SWOT mission archived at GHRC to complement hazardous weather related to floods caused by excess precipitation. This would complement other flood and extreme event (including precipitation) data sets.	Open. New Recommendation #7
20	GHRC should include GOES GLM data in its portfolio of accessible data, whether stored in house or as a virtual data set. Functionality should be seamless with other holdings..	Open. New Recommendation #8

New/Open Recommendations

New Recommendation #1 (previously Recommendation #3): GHRC should continue to hold AMS and AGU town halls, develop and distribute information brochures that describe their capabilities to potential data providers (e.g. field campaign PIs) and data users, utilize the NASA Hyperwall, and pursue other opportunities (BAMS) to enhance GHRC visibility.

Recommendation #3 (old; Table 1) was to attend several venues and increase GHRC visibility. GHRC staff attended a number of professional conferences and workshops, but it was unclear exactly what informational brochures or exact material was presented in town halls. (The annual report did communicate the new vision and mission of the DAAC.) Several UWG members suggested hands-on activities might be of value, particularly in the context of an AGU hyperwall/NASA booth. *GHRC seems to have a plan for AGU this year and should continue to have a presence at such venues on an annual basis. The UWG recommends establishing relationships with PIs and Program Managers to get access to invited program-level meetings and relevant data workshops.* Having an internal plan to do this would be an ongoing recommendation, but does not require any specific actions vis-a-vis the UWG.

New Recommendation #2 (merger of previous Recommendations #4, #5 and #6):

We feel that the GHRC has largely addressed the initial UWG recommendations from 2014-2015, which were broadly geared towards developing a unified theme for the DAAC, fulfilling that theme through current and planned data holdings, and developing tools that will better facilitate their users interaction with those datasets. The UWG feels that the next major issue for the GHRC is to continue this momentum by further developing their 5-10 year plan. The GHRC is in a unique position to be a data provider for “Storm Induced Hazards”, which already includes the following major categories: lightning, hurricanes and floods.

New Recommendation #2a: As part of their 5-10 year plan, the UWG recommends that GHRC should become the data subject matter expert on those categories — either through datasets that they provide in house or links to outside data — and become known within the community as the go-to place for datasets related to these subjects.

The UWG has decided to merge Recommendations #4, #5, and #6 into this New Recommendation, as responses to these recommendations should contribute to the 5-10 year plan.

New Recommendation #2b (previously Recommendation #4): Carry out dataset holdings analysis and create a reporting structure that categorizes what is available at GHRC and possibly elsewhere. This compilation should enable prioritization of efforts that will fill the most significant data voids, where these efforts align with the GHRC mission.

The first half of this recommendation from 2015 was closed. The second half of the recommendation was left open. Specifically, Recommendation #4 (2015) suggested that “it would be helpful to see data holdings and mission broken down into a simple chart describing the data holding inventory broken out by and related to GHRC mission components.” The chart (or

equivalent) would provide a mission-driven means to gap-fill the dataset inventory and assist in developing tactical and strategic planning related to dataset acquisition.

Working between the 2015 and 2016 UWG meetings, there has been more focused “bottom up” thinking on behalf of GHRC as it pertains to dataset inventory considerations (“what, who, where, why” approach) in the context of integrating data inventory with the GHRC mission. It is also clear that thought has been given to considering how GHRC could benefit from potential new dataset holdings in a strategic sense. For example, discussion of several potential new datasets discussed on day 1 “fit” the mission of GHRC and were presented in the context of mission when the question “why” was answered. This was especially true for tropical cyclone/convection focus where the bulk of the new datasets mentioned applied (e.g, SMAP Ocean Winds, TMI V7, TCIS, inter-calibrated TBs, CPEX field campaign). As another example, active pursuit of the SWOT dataset is ongoing and a good strategic move for supporting hydrologic hazards emphasis and synergies with ongoing precipitation dataset collections.

Hence there has been significant forward movement in addressing Recommendation #4, though it is not clear that the recommendation should be closed completely. *A chart (or equivalent) that clearly and succinctly demonstrated the linkages or traceability of current/proposed future data holdings to the GHRC mission components should be completed. Perhaps a hierarchical chart that showed traceability to specific GHRC mission components and then continued to NASA Earth Science focus area(s) would be useful. In addition to strategic planning such a chart would be useful for presentation to higher-level managers as well.* Recommendation 4 could be closed quickly when this “traceability” chart is completed.

New Recommendation #2c: Update public dataset information pages to include data holding analysis results that might be helpful to the user community (previously Recommendation #5). Determine a set of useful user metrics, with feedback obtained from the UWG, which can be routinely updated and made available to the NASA sponsor, UWG and broader community. Analysis of these metrics should inform the five-to 10-year plan (previously Recommendation #6).

Declared to be in progress for the time being, but not specifically addressed. It would have been nice to see a page on the GHRC site that simply displays user metrics, hot datasets, etc., but it is not obvious that this exists. Also, there was no mention of the Dashboard tool that was presented last year, which looked promising; however, there is a move towards using Google Analytics to analyze what parts of the site are being used, where they’re from, etc. which was presented by Deborah Smith.

New Recommendation #3 (previously Recommendation #8): Create a data lifecycle process for GHRC that can be applied to current and future holdings. Ask NSIDC and PODAAC for their policies and assess utility within GHRC. Publish the data lifecycle on the website, along with a contact, to provide clarity on the process for investigators interested in providing data.

The questionnaire, lifecycle plan and work flows have been completed and are now on the website. So this part of the recommendation has been met. However, the location of the publish data information is not intuitive.

The UWG recommends that lifecycle plans and workflows should be moved up a level (e.g under Resources instead of Resources>Documents)

New Recommendation #4 (previously Recommendation #11): Determine LIS technical specifications for data products, latency, formats, etc. Publicize this future data source at appropriate venues.

This recommendation addresses the future importance of LIS on ISS data to the GHRC, also emphasizing outreach with regard to the GHRCs lightning holdings. While LIS on ISS is an upcoming mission, the UWG would like to see the GHRC be more proactive about future data holdings, and this is a prime example. To close the recommendation, the UWG would like to see this future dataset publicized, and technical specifications of the data products and latency available to potential users before the start of the mission. The GHRC presentation claims that this effort is in progress, though not much ISS/LIS information can be found on GHRC websites. Therefore, the recommendation remains open.

The UWG envisions GHRC becoming a hub of all lightning data (i.e., New Recommendations #2 and #8). In addition to the existing lightning datasets at GHRC, they should also provide the services of virtual holdings, including brief introductions and links to all available lightning data from other resources, such as GLM data at CLASS, or other future lightning datasets from other countries.

Recommendation #5 (previously Recommendation #12): Develop a single tool that can provide broad use to multiple field campaigns and data types.

Significant progress has been made, but the scope of the effort needs to be better defined. The Python tools are very helpful to those that use Python, but other code for other languages should also be considered. The Field Campaign Explorer is very impressive, although work remains to be done to make it more useful for more users. For example, it would be useful to apply such a tool to other field programs serviced by the GHRC (e.g., the numerous GPM field campaign datasets it archives). The change from Oracle to Open database saves money and allows for improved flexibility, as well as normalized and simplified database schema. Hydro 2.0 is considerably more superior to previous version, and certainly appears greatly simplify accessing desired datasets.

While significant progress has been made, the UWG recommends that there is considerable work remaining on these tools, so this recommendation will remain open.

Recommendation #6 (previously Recommendation #17): Create Data bundles for scientists who want to study processes. Demonstrate such bundling capabilities for review by the UWG.

The committee was impressed by GHRC efforts towards bundling data for users who may be interested in processes. The virtual collection is a great effort and the case presented in Bugbee's talk on the GCPEX snow microphysics case study is an example of the GHRC addressing this recommendation. It appears to be a prototype for this type of data bundling. The micro-articles are another great way of offering the users an example of how to bring multiple data holdings together, however, it's not clear to the committee how this effort will be continued (i.e., who will determine the virtual collections in the future, and who will do the work? GHRC or the PIs?).

We recommend that the GHRC continue to pursue this effort, and in particular, feel free to use the UWG as a sounding board for micro-articles that the GHRC feels would be interesting for publication on the website. The UWG envisions a process by which the GRHC proposes a topic for a micro-article to the UWG Chair, who will then seek the advice from the subject-matter experts on the UWG. Also, the idea of using an automatic program to do data bundling needs more research. Giovanni could be a great tool, but lacks advanced logic in that it seems to have trouble identifying the most relevant datasets for a specific event. For these reasons, the recommendation remains open.

Recommendation #7 (previously Recommendation #19): Discuss the possibility of getting land data from SWOT mission at GHRC to complement hazardous weather related to floods caused by excess precipitation. This would complement other flood and extreme event (including precipitation) data sets.

Some progress has been made on this recommendation based on the presentation. GHRC has identified specific datasets from the SWOT mission (pass-based lake and river levels and 21-day averaged levels for each). In addition to the planned datasets, they have identified the field campaign mission, AirSWOT, and any value-added products from the early-adopter community as potential datasets. GHRC should make the pursuit of these dataset a priority during upcoming years, as they are relevant to this “storm induced hazard” aspect of GHRCs mission.

It is recommended that GHRC develop a plan to obtain the targeted SWOT datasets. Specific steps for the SWOT data holding plan can include expanding surface water hazard products and pursue existing flood products, such as the MODIS flood product, to develop a user base and reputation within the field. In addition to developing the product collection, GHRC should be in contact with PIs and program managers prior to the SWOT mission to begin discussions on the SWOT products.

Recommendation #8 (previously Recommendation #20): GHRC should include GOES GLM data in its portfolio of accessible data whether stored in house or as a virtual data set. Functionality should be seamless with other holdings.

GHRC should continue planning to be the first stop for accessing global lightning data. GLM should be managed as a virtual data collection since it is funded for data stewardship by NOAA. GHRC should plan to coordinate with NOAA CLASS/NCEI-NC and also maintain awareness of on-going coordination through the WMO of the upcoming operational lightning data from operational space agencies (e.g., refer to the SATURN Satellite User Readiness portal).

The GHRC has the only historic data archive for the NASA LIS data since launch of OrbComm-1 with OTD (1995-2000) and TRMM with LIS (1997-2015), as well as ancillary reference data from the commercial lightning data providers (e.g. Vaisala, Earth Networks, WWLLN, and select LMA regional data sets). Users can be found worldwide. These should be continued to be preserved with the launch of the GOES-R series that will provide an additional 20+ years of lightning data from the NOAA Geostationary Lightning Mapper for the western hemisphere, as well as the ISS-LIS data with planned launch in 2016 (also providing coverage to 54° latitude as does GLM). Also, China will host a Lightning Imager on their next series of FY-4 GEO satellites with first launch in December 2016, as will EUMETSAT with the launch of a Lightning Imager on their next series of MTG GEO satellites. The GLM, MTG-LI and CMA-GLI L1B and L2+ data (even, group, flash data components are in the same format as the LIS data structure). Thus, the new GEO satellites provide an extension of the initial LIS climate data set. The GLM L1B and L2+ are operational NOAA data products with long-term stewardship funded by NOAA and archived in the NOAA Comprehensive Large Array Storage System (CLASS) managed by the National Centers for Environmental Information, NCEI-NC). GHRC should coordinate with CLASS so the data are viewable at GHRC and point to CLASS, perhaps providing readers for GHRC user communities.

The GHRC also has field campaign holdings supporting field campaigns. *The reference data collected during the GOES-R Field Campaign should also be bundled with the satellite data for one-stop shopping.* The GOES-R Program is funding GHRC to develop a web portal for the 2017 GOES-R field campaign. NOAA CLASS is making plans for permanent stewardship after the campaign is completed. Again, coordination between GHRC and CLASS is desirable and should be a metric/milestone for FY17.

Closed Recommendations

Recommendation #9: Develop some guidelines or work flows for GHRC to handle future field campaign data. Online data questionnaire form should be developed further and shared with other DAACS.

The presentation given on Tuesday highlighting the various swim lanes and procedures for the automation of data demonstrated that the team has spent a lot of time considering how to improve their process for the collection and archival of field campaign data sets and associated metadata. The creation of their submission process is a great step towards putting more responsibility on the data set developer to provide the GHRC with information necessary to improve the automation of field campaign collections, which will improve overall efficiency and make it easier for them to achieve their mission in this area. The workflows that have been established seem well thought out and intentional, and they have demonstrated the utility of these workflows in the publication of data as a demonstration with expected testing and rollout in late 2016 and early 2017, respectively.

This recommendation should be considered closed/complete, *though the team should continue to use opportunities in future field campaigns to gather input from the community on how to further*

improve this process for science team members responsible for contributing data, and look for means of extending these capabilities to other DAACs and science teams, as appropriate.

Recommendation #10: Develop a data maturity model for GHRC data. Provide this on website and include maturity information for each dataset provided. Review NOAA's data maturity model (<http://www1.ncdc.noaa.gov/pub/data/sds/maturity-table-6level.pdf>) as a starting point.

A data maturity model has been implemented for GHRC data and is linked on the website. The maturity model is tied to levels of service. Datasets with higher levels of service are more closely evaluated for maturity. Lower levels of service are being less closely evaluated. As long as there has been some degree of evaluation for each dataset, the panel defers to the judgment of GHRC on the appropriate level of evaluation.

This recommendation should be considered closed (and was already denoted as Done in the GHRC presentation). The GHRC team has implemented a process by which to evaluate and assign maturity levels as a standard practice.

Recommendation #14: Communicate with the LPDAAC to understand their transition to HTTPS process. Provide highly visible examples, links to examples via email, and as much visibility as possible to ease the transition. A page with examples of different methods to download data, accompanied by example code, would be helpful.

Although they did not speak to this transition in this level of detail, it's obvious that this transition has occurred and therefore this recommendation is closed.

Recommendation #16: Explore and identify future users of possible mobile apps for NRT data. An assessment of how GHRC ingests format requirements could be used to broaden app utility.

The recommendation for a mobile app for NRT data is not considered a pressing issue for the GHRC to address, therefore the UWG will close this recommendation. GHRC has made an effort to make their website more mobile friendly, and some GHRC data holding may be accessible through Worldview on a mobile device.

Recommendation #18: Develop an attractive visualization that goes along with the new mission and vision statement that would help audiences associate the GHRC with its vision and mission statement.

This was accomplished through a 5.5 minute video. One UWG member suggested having an additional 30 seconds at the end showing a user going through the data download and usage process. *More such videos — perhaps shorter and targeting different audiences — would be valuable. We recommend having a general video on GHRC, then having a series of more detailed videos on specific field campaigns, online tools, etc.* Suggest Morgan Freeman to narrate. This recommendation is now closed.

New Member Suggestions for 2017 UWG

Table 3 provides a list of suggested new members for the GHRC UWG. There are two current members leaving (Table 1), although Chris Kummerow has offered to remain for one more year. A suggestion for the UWG would be to rework the “Expertise” categories as many of the members have overlapping expertise, but that will be worked on for next year’s annual meeting. Note that invitations for these suggested users have not gone out yet, but the UWG Chair and Co-Chair will reach out to them as soon as possible and well before the 2017 UWG meeting.

Name	Affiliation	Email
Courtney Schumacher	Texas A&M Univ.	cschu@tamu.edu
Pierre Kirstetter	U. Oklahoma	pierre.kirstetter@noaa.gov
Katrina Virts	NASA MSFC	kvirts@washington.edu
Russ Schumacher	CSU	russ.schumacher@colostate.edu
Scott Rudlosky	CICS/Univ. Maryland	scott.rudlosky@noaa.gov
Dan Cecil	NASA MSFC	daniel.j.cecil@nasa.gov
Wiebke Deierling	UCAR	deierlin@ucar.edu
Ian Giammanco	Insurance Institute for Business and Home Safety	igiammanco@ibhs.org