

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
+ Google Cloud Office Hours**
August 16, 2023 | 12-1pm

TO: NWS NODD Google Office Hours Participants

DATE: 16 August 2023 | 12-1pm EDT

FROM: Cindy Elsenheimer, Yuejian Zhu, Dr. Neil Barton, & Dr. Bing Fu (NOAA National Weather Service)
Adrienne Simonson, Jenny Dissen & Kate Szura (NOAA Open Data Dissemination Engagement and Communication)

SUBJECT: Responses to Questions from GEFS Office Hours



NOAA's National Weather Service (NWS) + Open Data Dissemination (NODD) + Google Cloud
Environmental Data Office Hours

August 16, 2023 | 12-1PM EDT | [REGISTER HERE](#)

- ❖ Share experiences on use and access of Global Ensemble Forecast System (GEFS) via Google Cloud
- ❖ Hear about data access via NOAA Open Data Dissemination (NODD)
- ❖ Connect with NOAA scientists, data leads, and data users

Adler Santos, Engineering Lead, Google Cloud Datasets

Adrienne Simonson, Patrick Keown, Jenny Dissen, Kate Szura, (NODD)

Cindy Elsenheimer, Partnership Engagement Lead, NOAA NWS Office of Organizational Excellence

Dr. Neil Barton, Physical Scientist, NOAA National Weather Service NCEP

Yuejian Zhu, Senior Meteorologist, NOAA National Weather Service NCEP

Dr. Bing Fu, Physical Scientist, NOAA National Weather Service NCEP

Dear Colleagues,

Thank you again for your tremendous contribution during the NWS NODD Google Office Hours. Your data related questions and comments raised during the discussion were heard and noted by NOAA.

This document provides brief responses to questions that were identified during the registration and that were raised during the discussion. Names and attributions of individuals and their affiliation have not been documented, unless it is a NOAA speaker.

We recognize the importance of continued engagement and collaboration, and invite ongoing comments via our emails.

Thank you,
Cindy, Yuejian, Neil, Bing (National Weather Service)
Adrienne, Jenny, Kate (NODD Engagement and Communication)

Contents for the Summary Notes

- [1. General Agenda of the Webinar](#)
- [2. Questions and Responses](#)
- [3. Office Hours Organizing Team](#)
- [4. Poll Results](#)
- [5. Resources / References](#)

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
+ Google Cloud Office Hours
August 16, 2023 | 12-1pm**

1. General Agenda of the Webinar

OUTLINE FOR THE DISCUSSION

- 12:00 - 12:05 Brief Introductions by NWS, NODD, and Google
- 12:05 - 12:25 Overview of Global Ensemble Forecast System (GEFS) & Access via Google
- 12:25 - 12:30 Invited Comments by NCICS and Users of GEFS via the Cloud
- 12:30 - 12:55 Open Discussion (Please use “Raise Hand” or the chat to raise questions)
- 12:55 - 1:00 Summary Comments/Closing Remarks/Next Steps

2. Questions and Responses

The questions below were identified as part of the registration process and during the Office Hours discussion. Responses are provided in brief where the NOAA team felt information was available.

QUESTIONS RAISED FROM REGISTRATION FORM

QUESTION FROM REGISTRATION	RESPONSE
<p>Dissemination of ensemble (EnKF.GDAS) restart data</p> <p>Reframed by NODD: Can you provide more information about the EnKF.GDAS restart data dissemination?</p>	<p>Thank you for the feedback. Additional information is needed to better understand the question.</p> <p>Some of the data is available via the GFS buckets. NOAA recognizes that the latency is not optimal, and may not include all restart files for the ensemble. NOAA also realizes that the very high data volumes are a challenge.</p> <p>NODD has started internal discussions on alternate methods for internal access, however, on-going data transfers of that volume directly from our super computer system remains a challenge. We will continue to discuss this with our colleagues.</p> <p>We welcome input and feedback from the user community on the use case, the need, and the importance of the value that this data provides.</p> <p>Below are links to the NOAA NCEI systems: NCEI Link to GDAS product information Access to GDAS</p>

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
+ Google Cloud Office Hours
August 16, 2023 | 12-1pm**

<p>Open data</p> <p>Reframed by NODD: Can you provide more information about the NOAA Open Data program?</p>	<p>GEFS is available via the NOAA Open Data Dissemination (NODD) program efforts, via all 3 cloud partners.</p> <p>Please visit the NODD website at noaa.gov/nodd/datasets for a list of datasets available via the cloud, and how to access those datasets.</p>
<p>Timely use of operational GEFS data for ocean forecasting - we are tired of NOMADS and are eager to access the data in other ways</p>	<p>Cloud access provides an alternative way to access GEFS data.</p> <p>Access via Microsoft Planetary Computer: https://planetarycomputer.microsoft.com/dataset/storage/noaa-gefs</p> <p>Access via AWS RODA is here: https://registry.opendata.aws/noaa-gefs/</p> <p>Access via Google Cloud: https://console.cloud.google.com/marketplace/product/noaa-public/gfs-ensemble-forecast-system?pli=1&project=valued-ceiling-316301</p> <p>gs://gfs-ensemble-forecast-system https://console.cloud.google.com/storage/browser/gfs-ensemble-forecast-system</p> <p>gs://global-ensemble-forecast-system-retrospective https://console.cloud.google.com/storage/browser/global-ensemble-forecast-system-retrospective</p>
<p>Could certain datasets available on AWS/GCP also be made available on Azure? (specifically, NWM?)</p>	<p>NODD currently makes available ~37+PB of NOAA data via the cloud partners. Please visit the NODD website at noaa.gov/nodd/datasets for details on access to the data.</p> <p>The National Water Model is currently available on Azure. It can be accessed on Azure Planetary Computer and Microsoft AI for Earth Datasets.</p>
<p>Authentication, big data downloads over browser (or not), training and tutorial materials for new users, streaming data scope</p>	<p>NODD understands this request and has shared it with the NOAA Cloud Program Office Leadership as well as the NWS Partners Lead, Cindy, who has shared it with NWS leadership.</p>

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
+ Google Cloud Office Hours
August 16, 2023 | 12-1pm**

NODD product entry set up	Thank you for the feedback. Additional information is needed to better understand the question.
Internet connectivity for NOAA	<p>Thank you for the feedback. Additional information is needed to better understand the question.</p> <p>NODD and NWS recognizes this concern and has logged this feedback for awareness for NOAA, NODD, and NWS leadership.</p>
Visualization of Data	Thank you for the feedback. Additional information is needed to better understand the question.
GEFS/EnKF-GDAS restart data on AWS S3	<p>NODD understands this request and will discuss it internally and with the NODD cloud partner.</p> <p>NODD makes available GEFS via the AWS Registry of Open Data. The link for AWS RODA page is https://registry.opendata.aws/noaa-gefs/.</p>
<p>Access to all files needed to run GEFS</p> <p>(Reframed by NODD: Where and how can the users have access to all files needed to run GEFS?)</p>	<p>The initial data under init directory should be sufficient for running GEFS.</p> <p>https://noaa-gefs-pds.s3.amazonaws.com/index.html#gefs.20230815/00/atmos/init/</p>
GEFS Wave Station File Data Discrepancies (Specifically spread values are inconsistent with gridded data values)	Additional information is needed to better understand the question. Information about which buoy point and a date would be very informative to see an example and understand if this is a widespread issue.
Next gen met system	Thank you for the feedback. Additional information is needed to better understand the question.
<p>Is there any significance/rationale to certain data being stored in a certain CSP (AWS vs Google)?</p> <p>What will the processes for loading and maintaining metadata for the CSP landing pages look like?</p>	<p>The first and most important rationale is the need of the NOAA data owner for data dissemination. Then second and equally important are the needs of the users. We take user requests into consideration when adding data to certain CSPs.</p> <p>Once data is loaded onto a CSP, we work with the respective CSPs to upload information to generate the landing pages. This process is different for each CSP as well as the efficiency of getting the information uploaded.</p>

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
+ Google Cloud Office Hours
August 16, 2023 | 12-1pm**

	<p>NODD welcomes feedback from the user community on datasets of interest they'd like to see in the cloud. Please email us at NODD@NOAA.GOV</p>
<p>An overview of how to access GEFS data via the cloud and a case example would be great.</p>	<p>This Office Hours provided an overview of how to access GEFS via Google Cloud. Please send us an email at NODD@NOAA.GOV for requests related to access via Microsoft and AWS.</p> <p>Access via Microsoft Planetary Computer: https://planetarycomputer.microsoft.com/dataset/storage/noaa-gefs</p> <p>Access via AWS RODA is here: https://registry.opendata.aws/noaa-gefs/</p> <p>Access via Google Cloud: https://console.cloud.google.com/marketplace/product/noaa-public/gfs-ensemble-forecast-system?pli=1&project=valued-ceiling-316301</p> <p>gs://gfs-ensemble-forecast-system https://console.cloud.google.com/storage/browser/gfs-ensemble-forecast-system</p> <p>gs://global-ensemble-forecast-system-retrospective https://console.cloud.google.com/storage/browser/global-ensemble-forecast-system-retrospective</p>

QUESTIONS / DISCUSSION FROM THE OFFICE HOURS

QUESTION FROM DISCUSSION	RESPONSE
<p>Is EnKF an evolving state for initial conditions? It couldn't be reproduced algorithmically?</p>	<p>EnKF is a method to determine how much the initial conditions should differ between ensemble members.</p>
<p>Is it a function of previous simulations, or could perturbations be generated on demand?</p>	<p>Yes, it's a function of the previous simulations in the data assimilation cycle.</p>
<p>Will these slides be shared after this presentation? Very nice to</p>	<p>Slides and notes from the discussion will be available and posted on the NODD website and the NWS website. Links, notes, and the slides will also be shared with</p>

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
+ Google Cloud Office Hours
August 16, 2023 | 12-1pm**

<p>have this information in one place :)</p>	<p>participants. NODD website: www.noaa.gov/nodd NWS Calendar Website: https://www.weather.gov/wrn/calendar</p>
<p>Is GEFS available on GCP or just AWS at the moment?</p>	<p>GEFS is on NOMADS and on Google Cloud Platform. You can access GEFS via Google Cloud at https://console.cloud.google.com/storage/browser/gfs-ensemble-forecast-system or by navigating to the marketplace page https://console.cloud.google.com/marketplace/product/noaa-public/gfs-ensemble-forecast-system?pli=1&project=valued-ceiling-316301</p>
<p>Does the ensemble member carry to the next day? Does ensemble member 1 today have any relation with ensemble member 1 yesterday or ensemble member 1 tomorrow?</p>	<p>The ensemble member 1 today has no relation with ensemble member 1 from yesterday or tomorrow, they are independent.</p>
<p>How is precipitation phase computed? Temperature threshold, or is humidity used?</p>	<p>If you are asking about precipitation type. A “energy method” is being used to determine precipitation type in upp. https://github.com/NOAA-EMC/UPP/blob/develop/sorc/ncep_post.fd/CALWXT_BOURG.f You can check Bourgoiuin’s 2000 paper: https://journals.ametsoc.org/view/journals/wefo/15/5/1520-0434_2000_015_0583_amtdpt_2_0_co_2.xml Precipitation type (rain, snow, graupel) is based on temperature, moisture, and potentially other variables depending on code used within the model. This code is generally updated frequently.</p>
<p>What drives which variables are in the pgrb2sp25 downscaled file output? Could other variables potentially be added?</p>	<p>The variables in pgrb2sp25 are determined by “gfs_pgrb2a_f00.parm_0p25” and “gfs_pgrb2a_fhh.parm_0p25”. Yes, other variables can be added as long as the variables are within the list of variables from upp.</p>
<p>Are the static directories like fix_20200927 or fix_nco_gefsv12.3</p>	<p>The UFS global-workflow staff mentioned that it will be shared in the cloud in the future.</p>

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
 + Google Cloud Office Hours
 August 16, 2023 | 12-1pm**

<p>shared somewhere? In other words, inputs to the simulator as opposed to outputs?</p>	
<p>How do these differ from the GEFS Restarts stored in AWS? https://noaa-gefs-pds.s3.amazonaws.com/index.html</p>	<p>The GEFS restarts stored on AWS (and other cloud platforms) are for a particular date to start the model. The “fix” files are fixed/static in time and the same data are used for each date the model is started</p>
<p>Is the spread calculated the same GEFS Wave Stations Output as it is for the grib output? In my attempts to use it, the spread values do not match up with the grib output, but the mean values do.</p> <p>ex: https://storage.googleapis.com/gefs-ensemble-forecast-system/gefs.20230522/18/wave/station/gefs.wave.t18z.bull_tar</p> <p>I then extracted the archive with tar and opened the gefs.wave.44097.bull file in a text editor which reports the Hs spr variable as 6.53 meters, which is a very large value for the standard deviation when Hs avg is only 1.08 meters. All of the stations seem to have the same issue so not sure if there is documentation on why this is widespread, all stations output files have spread values that do not match their grib counterparts</p>	<p>Thank you for the feedback and for providing this information. NODD and NWS note this mismatch and have logged this feedback for awareness for NOAA, NODD, and NWS leadership.</p>
<p>Is this a companion or a replacement for accessing GEFS via NOMADS?</p>	<p>At this point, cloud access is a companion. NOMADS is more than just a webservice that makes binary data available, and we are not replicating or replacing any of that. If you are using services such as Google, you can</p>

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
+ Google Cloud Office Hours
August 16, 2023 | 12-1pm**

	spin up the workload for extremely fast data access without having to download the data.
Do you gate or throttle any of the data accessed from the cloud providers?	No, NODD does not throttle data. Google Cloud Storage has unlimited bandwidth. Should your local bandwidth be the challenge, consider using the multiprocessing flag to parallel download. Google does have a limit API, so it may give you “slow down” messages. But those aren’t specific to this dataset, just Google service rate limits.
Where does the two-day delay for HPSS come from? Can NODD pull from production systems?	We have access to the research system for this activity, for other things we may have access to the other system, but it takes up to two days for this data to be available on the research system.
When data gets synced to the cloud provider, when doing backfills to correct erroneous data, notifications get sent out?	If this is a problem for your system, we suggest backfilling by date or age.
Where can you get access to the fixed and static datasets required to run global ensemble forecasts? Namely I can't find these anywhere: /gpfs/hps3/emc/ensemble/noscrub/emc.enspara/common/git/fv3gefs/fix_20200927 /scratch2/NCEPDEV/ensemble/noscrub/common/FIX/gefs/fix_nco_gefsv12.3	We have not been able to onboard this data onto NODD, but it is an open task that we are tracking. If we can find a way to get access, we will make it possible. However, if we pull it from the research side, it will have a 2-day delay. Those data are only available on NOAA WCOSS or R&D machines, but we are planning to share it in the cloud.
Are you able to provide a roadmap of additional services, parameters, and data you plan to provide?	The user community drives the priority. We want to understand the demand for need and asks, and this drives how we can make the data available and accessible. We work with many stakeholders and cloud partners to determine the data that can be made available.

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
 + Google Cloud Office Hours
 August 16, 2023 | 12-1pm**

Is GEFS currently running on the cloud? Any plans?	Not at this moment. There is currently no plan to do so and it will take some infrastructure planning. It is possible to run the seasonal forecast system on the cloud in the future, which will have a similar structure to the GEFS.
--	--

3. Office Hours Organizing Team

Name	Title
Cindy Elsenheimer	NWS Partner Engagement Lead
Yuejian Zhu	NWS Senior Meteorologist
Dr. Neil Barton	Physical Scientist
Dr. Bing Fu	Physical Scientist
Adrienne Simonson	NODD Director
Patrick Keown	NODD Program Manager
Jenny Dissen	NODD Engagement Lead / NCICS / NC State University
Katelyn Szura	NODD Communications Lead
Jonathan Brannock	NODD Lead Cloud Software Engineer /NC State University
Otis Brown	Director, NC Institute for Climate Studies (NCICS) / NC State University
Adler Santos	Engineering Lead for Google Cloud Datasets

**NOAA National Weather Service (NWS) + NOAA Open Data Dissemination (NODD)
 + Google Cloud Office Hours
 August 16, 2023 | 12-1pm**

4. Poll Results

Poll 1		
Question	Answer	Count
How do you access GEFS data today?	On-prem via NOAA	5
	Cloud	4
	Both/Either	2
	3rd Party/Web-based Viewer	2
	None/Other	16

Poll 2		
Question	Answer	Count
My primary goal for attending today is:	Technical use and access of GEFS data	5
	To learn about cloud access to date (e.g. NODD Program)	17
	Meet and engage with NOAA staff scientists	3
	Learn about Google Cloud access and tools	7

5. Resources / References

- [NOAA Open Data Dissemination | NODD Email](#)
- [NWS Office of Organizational Excellence | Cindy.Elsenheimer@NOAA.GOV](#)

Thank you to our participants for engaging in this discussion!