# Toward a petascale data hackathon for exploring a digital twin of the earth

*Valentine Anantharaj <vga@ornl.gov>, Suzanne Parete-Koon, and Thomas Papatheodore*
Oak Ridge National Laboratory

OAK RIDGE National Laboratory | OAK RIDGE LEADERSHIP COMPUTING FACILITY

## Background

During 2020-2021, the European Center for Medium-Range Weather Forecasts (ECMWF) and ORNL received an ASCR INCITE award[1] for global 1-km climate simulations, called a nature run (NR), using ECMWF's Integrated Forecast System (IFS).

The project team, led by Nils Wedi, completed the first set of simulations in just under 5 months, using Summit at the OLCF. The team analyzed 250 TB of data over the next three months, and published their first manuscript[2].

The NR simulations revealed unprecedented detail of the earth's atmosphere (Fig 1), defining a baseline for a digital twin of the earth for weather, climate and energy applications (Fig 2), including extreme events.

Then, OLCF started getting requests for our data from the research community. Distributing the data became a logistical challenge. In addition, the community also requested case studies of extreme events, such as a tropic cyclone (TC) and three severe weather events over the continental US.

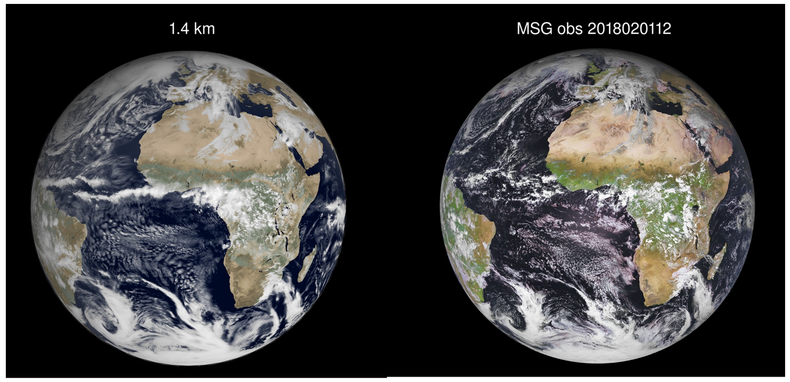


**Figure 2:** A baseline for a digital twin of the earth: Simulated satellite view (left) and observation from a satellite (Courtesy: P. Lopez, ECMWF).

## Challenges

The simulation output was stored as spherical harmonic coefficients. Extensions to the GRIB standard formats became a problem for many of the common tools used by the user community.

The users could not use the data in the native output forma. So, we remapped the spectral coefficients into grid point values inflating the data volume to 3 PB instead of 750 TB for two seasons and the four case studies.

Network bandwidth: even if the users found the necessary storage. It would take 30 days to transfer 1 PB of data at 3 Gbps via ESnet.

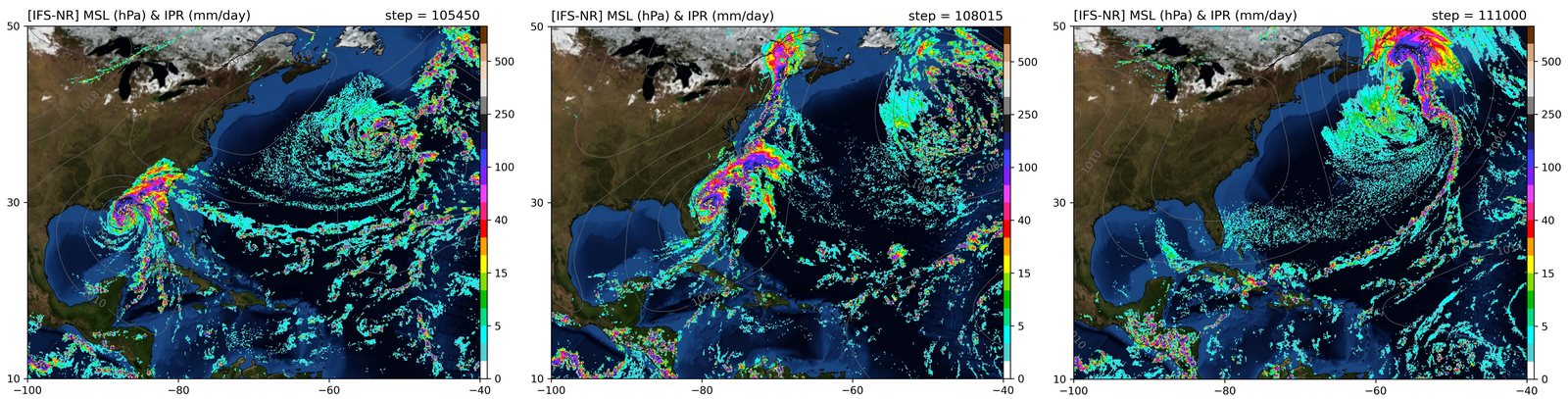


**Figure 1:** Spontaneously generated tropical cyclone (TC) in the 1km atmosphere-only simulations on Summit. Mean sea level pressure [hPa] is shown in contours and shading shows instantaneous precipitation rate in [mm/day]. TC hitting the west coast of Florida (left); passage of TC from Florida to the North Atlantic (right); and merging of the TC with the extratropical cyclone in the North Atlantic (right).

## Data hackathon

Three years ago, OLCF had anticipated such data-centric user needs for science, and envisioned a new type of project where the user community could bring their investigations to the user facility, and work with the data locally at OLCF rather than transferring large volumes of data to the host institutions.

We have formulated a data hackathon to facilitate access to the data. The user community is engaged in the planning process, involving potential users from national labs and universities – both US and international.

The formal announcement will be released at the end of May. We have requested 25,000 node-hours on Andes, the analytics cluster and 3 PB of storage on center-wide file system. In addition, the hackathon projects will also have periodic access to Ascent to meet AI/ML needs. We expect to select 5 - 10 projects with 2 to 4 users per project.

Users are interested in a diverse range of investigations using the data: observing system simulation experiments to plan for satellite instruments to monitor tropical cyclones and severe storms, understanding gravity waves in the atmosphere, local and mesoscale weather phenomenon, etc.

The hackathon will be launched next month, in June 2022.

## References and Acknowledgments

[1] Wedi, N., Dueben, P., Bauer, P., and Anantharaj, V (2020-2021). A baseline for global weather and climate simulations at 1km resolution. DOE ASCR INCITE Award (2020-2021). https://www.doeleadershipcomputing.org/wp-content/uploads/2020INCITEFactSheets_rev.pdf

[2] Wedi, N. P., Polichtchouk, I., Dueben, P., Anantharaj, V. G., Bauer, P., Boussetta, S., et al. (2020). A baseline for global weather and climate simulations at 1 km resolution. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002192. https://doi.org/10.1029/2020MS002192

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. ECMWF also benefited from collaborations funded via ESCAPE-2 (No. 800897), MAESTRO (No. 801101), EuroEXA (No. 754337), and ESiWACE-2 (No. 823988) projects funded by the European Union's Horizon 2020 future and emerging technologies and the research and innovation programmes.