THE PACE DATA HACKWEEK LAUNCHES A NEW WAVE OF OPEN SCIENCE

HOW WE DID IT (TWICE), AND WHY

By Anna E. Windle, Ian Carroll, Carina Poulin, Skye Caplan, and P. Jeremy Werdell

ABSTRACT. PACE Data Hackweeks bring together student, early-career, and senior scientists to work hands-on with data from NASA's Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission. During two past events, participants collaborated in small teams to explore scientific questions, develop reproducible workflows, and gain experience working with large, cloud-hosted datasets. Leveraging community-supported cloud infrastructure, the hackweeks lowered technical barriers and emphasized modern tools, open-source practices, and FAIR data principles. The events fostered interdisciplinary collaboration and led to tangible outcomes, including new analyses, shared code, and ongoing collaborations. Overall, the PACE Data Hackweeks demonstrate that mission-focused hackweeks are an effective way to build community capacity and support adaptation to evolving data and computing environments.

NASA DOES HACKWEEKS

The NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) Data Hackweek builds on a growing ecosystem of geoscience hackweeks that leverage NASA Earth Observing System datasets, including Geohackweek, OceanHackWeek, and mission-specific events like ICESat-2 and SnowEx hackweeks (Huppenkothen et al., 2018; Mitchell et al., 2026). Unlike traditional workshops or competitive hackathons, hackweeks emphasize participant-driven learning through a mix of coding tutorials and group projects in a collaborative setting. Demand for these events has grown alongside major shifts in Earth science data access, including the migration of petabyte-scale data to NASA's Earthdata Cloud, hosted on Amazon Web Services (AWS), and the launch of new satellite missions that introduce novel, complex data products. Launched in February 2024, the PACE mission carries a three-instrument payload that collects synergistic oceanic, atmospheric, and terrestrial observations that extend Earth system data records for the next generation of scientists. Details on PACE data file formats, types, and volumes can be found at the Ocean Biology Distributed Active Archive System (OB.DAAC) and PACE websites. The OB.DAAC provides processed data along with publicly available processing and visualization tools, enabling researchers to quickly engage in exploring applications and interpretations of PACE data products. However, the user community still faces the dual challenge of becoming familiar with new data products while adapting workflows to cloud-based computing environments, including the use of community-supported software and cloud infrastructure.

Hackweeks provide an effective mechanism for helping researchers adapt to these evolving data environments by promoting open-source tools, FAIR (findable, accessible, interoperable, reusable) data practices, reproducible workflows, and cloud-based collaboration in support of NASA's open science goals (Scheick et al., 2026). Here, we describe the goals and outcomes of the PACE Data Hackweeks, highlight participant experiences, and identify remaining challenges. We conclude that sustained investment in hackweeks and community-maintained cloud infrastructure is essential for maximizing scientific and applied value of NASA's evolving Earth science systems.

GOALS OF A PACE DATA HACKWEEK

From organizing and leading two successful in-person hackweeks, we have gained valuable insights and refined a suite of goals for teams interested in hosting similar events. During early planning stages, our team consulted with the ICESat-2 Hackweek organizers, whose experience and resources proved instrumental in shaping our approach. Members of our team have also participated in and organized a multitude of short-format trainings, some within the Carpentries or eScience Institute hackweek frameworks. These experiences allowed us to formulate the goals below as a high-level strategy for achieving a research-focused event that generates tangible scientific outputs.

GOAL 1: KICKSTART NOVEL RESEARCH AND NEW COLLABORATIONS

PACE Data Hackweek participants arrive with diverse research ideas and are expected to depart with well-defined projects, collaborative teams, and a foundation for sustained remote collaboration. Developing effective group projects is one of the primary organizational challenges, requiring early alignment of research goals and active guidance from both scientific and technical mentors. Daily science lectures and tutorials, modeled after the ICESat-2 Hackweek, helped shape projects while providing

exposure to current PACE research across oceanic, atmospheric, and terrestrial domains, as well as a break from intensive coding.

To promote continuity beyond the hackweek, teams were encouraged to develop projects into conference presentations, manuscripts, or proposal concepts, with continued support through cloud computing resources, Slack communication, technical guidance, and funding and publication advice. Structured social activities were also integral to fostering collaboration, helping participants build networks, exchange feedback across teams, and sustain a balanced, collaborative environment.

GOAL 2: DEMONSTRATE THE VALUE AND LIMITS OF CLOUD COMPUTING

The transition of NASA's Earthdata archive to the commercial cloud encourages researchers to adopt cloud-based data analysis (Gentemann et al., 2021). Earthdata Cloud provides free access to datasets stored in AWS and discoverable through NASA's Common Metadata Repository, but it does not currently offer compute resources; however, efforts are underway to expand access to integrated cloud-based computing environments that would allow users to analyze data without needing to manage cloud infrastructure. Cloud-based analysis requires separate compute infrastructure, most commonly AWS Elastic Compute Cloud (EC2). This architecture supports high-throughput, data-intensive workflows by enabling direct access to large datasets within the same AWS region, reducing data movement and speeding up processing.

An emerging model for accessing EC2 instances is through community-maintained cloud-based JupyterHubs. Co-locating hackweek compute and data in the cloud lowers barriers to working with large datasets without duplication, supports data streaming and reproducible workflows, enables real-time collaboration, and allows users to scale seamlessly from simplified tutorials to full-mission workflows (Box 1; Martin et al., 2025). For the PACE Data Hackweek, we used CryoCloud, a JupyterHub maintained by a broad Earth science community, and we collected participation feedback on usability during the event (online supplementary Box S1).

GOAL 3: GROW A PERSISTENT ONLINE RESOURCE

To extend the reach of the hackweeks, we created a public website using an open-source website template and the GitHub Pages hosting service. The PACE Data Hackweek website includes a single-page landing site for logistics and a detailed JupyterBook with recorded lectures, coding tutorials, schedules, and preparation materials. Participants found it intuitive and helpful, reducing pre-event questions.

Hackweek content is preserved as static archives and incorporated into NASA's Ocean Ecology Lab Help Hub, which provides

BOX 1. Summary of the value and limitations of using a cloud-based community-maintained JupyterHub for the PACE Data Hackweek.

+ VALUE	– LIMITATIONS
MANAGED COMPUTE ENVIRONMENT	
• Eliminates the need for individual software configuration, allowing participants to begin productive work immediately • Can be reused on other local or remote platforms beyond the hackweek • Reduces technical barriers and inequalities among participants • Fosters a community of support and collaboration	• Restricted control over resources and system configuration • Limited available memory (Box S3) • Unfamiliar development environment and operating system (i.e., JupyterLab, Linux, S3 bucket storage and access, commonly owned "shared" file system)
DIRECT DATA ACCESS	
• Reduces latency through storage-adjacent compute resources • Supports collaborative workflows with high-bandwidth connections and unlimited concurrency • Removes dependencies on local storage or network limitations	• Lack of cloud optimization for data products stored as NetCDF4/HDF5 reduces performance • Data products from different agencies (e.g., NASA, USGS, NOAA, ESA) are not all co-located in the same commercial cloud
FACILITATED COLLABORATION	
• Encourages reproducible research by standardizing the computing environment with access to datasets • Supports generation of preliminary analyses and hypothesis testing that can be built into future research proposals • Promotes sharing of code, workflows, and results with the broader community, accelerating scientific discovery	• Reproducing the compute environment outside a managed JupyterHub requires additional training • Limits and "best effort" restrictions on per-user and shared file storage (Figure S4)
COMMERCIAL CLOUD SCALABILITY	
• Supports both simple tutorials and computationally intensive projects • Enables seamless scaling for larger datasets or complex analyses	• Reliance on S3 over local or network file systems for persistent storage is unfamiliar and less widely supported

reproducible Jupyter Notebook tutorials combining code, text, and visualizations. Tutorials are collaboratively developed, peer-reviewed, and regularly updated using hackweek feedback to ensure relevance and usability. Hackweeks play a critical role in motivating the continual updating of Help Hub tutorials by serving as development sandboxes where project activities and participant feedback feeds directly into new or improved notebooks. Since its launch, the Help Hub has grown from 48 to over 6,200 weekly page views, with usage spikes following trainings and events (Figure S1).

GOAL 4: LEVERAGE PRIOR EXPERTISE

Hackweeks, workshops, and short courses are common in academic and open-source communities, often implicitly adopting the Openscapes program's flywheel approach: leveraging momentum toward greater outcomes (Robinson and Lowndes, 2022). Organizing a hackweek raises many questions about costs and execution, and planners are encouraged to consult multiple past event managers for guidance, as it is unlikely that any one approach can be exactly replicated. The PACE Data Hackweeks draw on two key sources of momentum: the CryoCloud community and the Ocean Carbon and Biogeochemistry (OCB) Project Office.

Participants receive access to CryoCloud (Snow et al., 2022), a cloud-based JupyterHub managed by the International Interactive Computing Collaboration (2i2c), originally developed for NASA cryosphere research but now serving a broader Earth science community. To reflect this expansion, the platform will soon be rebranded (Tasha Snow, Earth System Science Interdisciplinary Center, University of Maryland and NASA Goddard Space Flight Center, *pers. comm.*, 2025). Access during and after the hackweek is supported by NASA funding and AWS cloud credits.

Professional logistics support from the OCB Project Office assisted with funding, contracts, and travel, allowing the core organizing team to focus on participant recruitment, instructional design, and tutorial development. Other hackweeks have successfully partnered with the University of Washington eScience Institute for similar logistical coordination.

We also built on prior expertise by empowering past participants to take on leadership roles in subsequent events. These mentors brought valuable participant perspectives and offered practical feedback that helped improve the overall participant experience.

EXPERIENCE AT A PACE DATA HACKWEEK

The PACE Data Hackweeks, held in August 2024 and 2025, aimed to broaden access to PACE data, enhance data-access tools, promote open-source workflows, and support integration with other Earth observation datasets. Each five-day, in-person workshop brought together 40–50 participants, with all instructional materials released and new code and analyses shared across the community. Project teams were first organized around research interests identified in participants' applications and then grouped into broad themes. Mentors worked with groups to refine project scope, align goals with available datasets, and ensure that proposed analyses were feasible within the week-long program. This structure enabled teams to quickly coalesce around focused research questions while remaining flexible to emerging ideas. Through these collaborative efforts, participants explored novel applications of PACE data, developed reproducible workflows, and generated preliminary analyses that seeded new collaborations, informed future research directions, and in some cases, supported early-stage proposal development.

Of the 88 total participants across both years, 36 completed a post-event feedback survey and a separate demographic survey (41% response rate). Highlights from the surveys are included in the following summary of the hackweek experience and in the online supplementary material (Boxes S1–S3, Figures S2–S4). We also show data on cloud computing resource usage and costs that were derived from EC2 activity logs during the 2025 event.

LOGISTICS

Each five-day, in-person event was held at the University of Maryland, Baltimore County (UMBC) and funded through OCB Training Activity grants. OCB covered travel, lodging, and meals for US-based participants, while international participants received partial travel stipends, with remaining expenses supported by the participants or their home institutions. Accommodations were provided in UMBC dormitories, and meals were served at the university cafeteria. To support participants with family or dependent care responsibilities, additional funding was available to offset caregiver costs or to allow for customized travel arrangements.

SCHEDULE

Each day combined scientific lectures with hands-on coding tutorials and demonstrations designed to build skills in data accession, Python programming, and cloud computing (Figure 1). Both science lectures and coding tutorials were highly rated (based on 1–5 ranking scores collected in feedback surveys), and 100% of survey respondents reported increased confidence in working with PACE data (Figure S2). Daily check-ins supported collaboration and progress, and on the final day teams presented their projects and shared code via public GitHub repositories, with most planning continued collaboration (Box S2). Feedback was overwhelmingly positive, with participants describing the hackweek as fun, engaging, and collaborative (Figure S3). Social events, including a Bob Ross paint night, a fireside chat with NASA program managers, a local crab feast, a PACE Nerd Night, and a themed trivia exercise, were frequently cited as essential for community building and sustained engagement:

> "Best part of the day! Every night was so great and the energy was great."

> "I loved the intentional inclusion of social activities. It helped people meet and have fun while maintaining work-life balance and camaraderie."

DEMOGRAPHICS

The respondents represented a diverse international community spanning 15 countries (Figure 2). Most were graduate students or early career researchers, with the majority holding master's or doctoral degrees (Figure 2). Notably, 74% reported that the PACE Data Hackweek was their first hackweek experience.

CODING ENVIRONMENT

PACE Data Hackweek participants onboarded to CryoCloud benefited from a fully configured, cloud-based software environment with direct access to all PACE datasets and no local setup required. The hub launched Docker containers, running on adjustable hardware configurations (2–29 GB RAM and 0.2– 4 CPUs), that

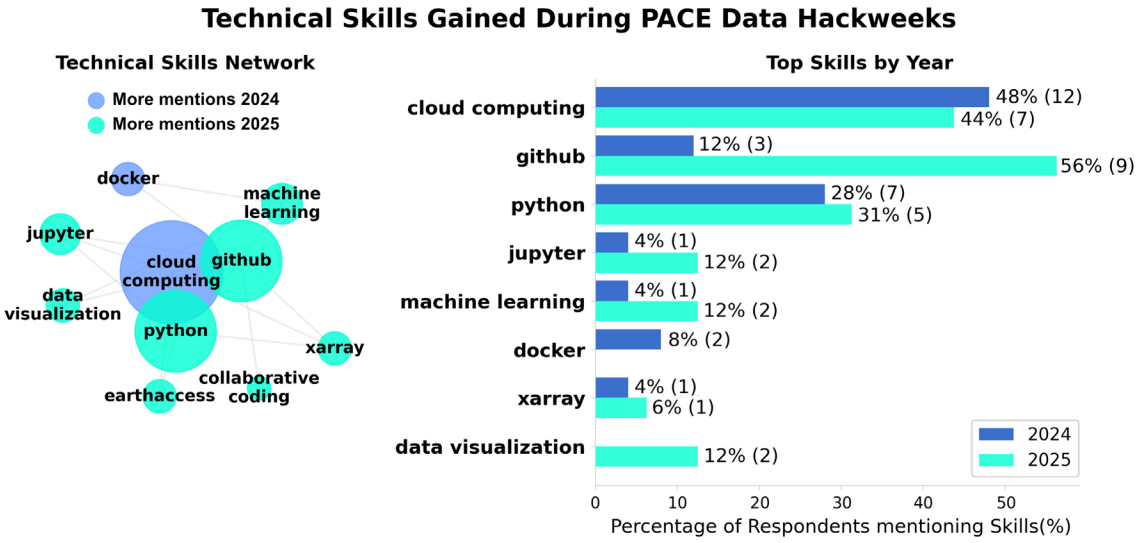

FIGURE 1. Most common technical skills, and associations among technical skills, mentioned in responses to the feedback survey question "What technical skills did you gain this week?" Node size in the technical skills network represents the number of times a skill was mentioned, while proximity indicates how frequently skills were co-mentioned by the same participant. Node coloring is based on the year in which the skill was mentioned most often.

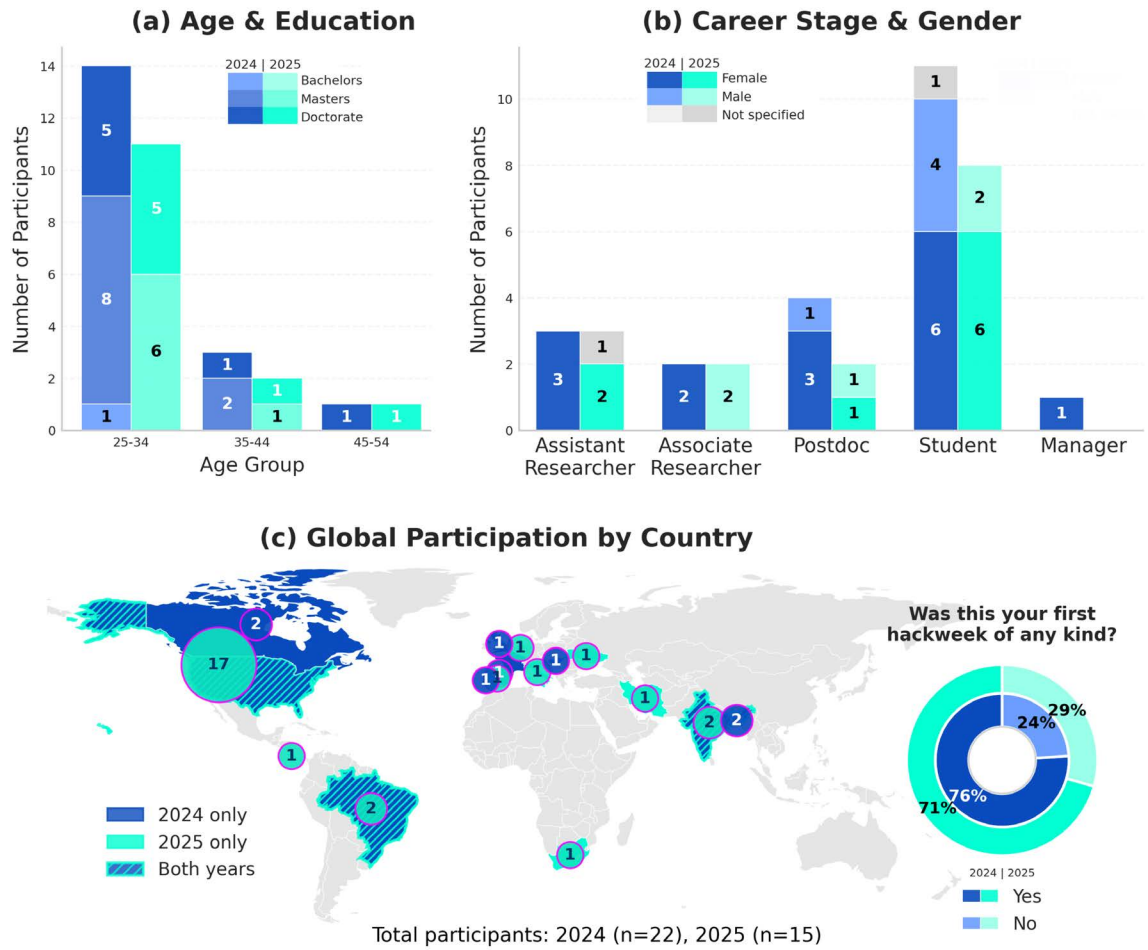

FIGURE 2. Responses to the demographic survey showing (a) age and education level, (b) career stage and gender, and (c) country of residence and first-time hackweek participation. Providing gender was optional and open-ended, with a non-response tagged as "Not specified."

included required Python packages and the NASA Ocean Biology Processing Group's Ocean Color Science SoftWare (OCSSW); this container is available on CryoCloud as the "Ocean Biology Python Image" (Figure S5). Participants were able to begin running tutorials and regenerating PACE data products in the commercial cloud almost immediately, all through a web browser.

Using the community-maintained JupyterHub, participants could execute tutorial and group-project notebooks, while experiencing both the benefits and limitations of cloud-based workflows (Box 1). Overall, CryoCloud played a central role in reducing technical barriers, fostering collaboration, and supporting open, reproducible workflows (Box S1). Feedback indicated that the majority of participants intend to continue using cloud computing in their research, although some concerns remain (Box S3, Figure S4).

COMMERCIAL CLOUD USAGE AND COSTS

Using Grafana dashboards developed by 2i2c, we charted EC2 usage and costs for all participants during the 2025 PACE Data Hackweek (Figure 3). Daily memory usage per user ranged from 155 KB to 29.4 GB, with an average of 1.79 GB, while CPU usage peaked at ~3.7 CPUs per user and averaged 0.6 CPUs per day (Figure 3). Peak memory and CPU demand occurred late in the week during intensive group project development.

CryoCloud costs include user-dependent expenses (e.g., compute and personal storage) and user-independent infrastructure costs that run continuously. Compared to the previous model of launching a dedicated hackweek JupyterHub that incurred ~$2,250 in startup and ~$2,250 in monthly infrastructure costs (Tasha Snow, Earth System Science Interdisciplinary Center, University of Maryland and NASA Goddard Space Flight Center, *pers. comm.*, 2025), maintaining a shared hub is substantially more cost-effective. For the full week, total EC2 compute costs were $367.99, with $36.36 in individual storage and $19.49 in shared storage, for a combined total of $423.84. Individual daily costs peaked on Day 4, with the highest single-user daily cost at $7.31 and a maximum single-user weekly cost of $22.61 (Figure 3). Uncertainty around these costs was a commonly raised concern among participants (Box S3), who did not have real-time access to the Grafana dashboards.

OPPORTUNITIES FOR IMPROVEMENT

While cloud computing is essential for processing large PACE datasets and enabling hackweeks, several challenges remain. (1) CryoCloud provides user home directories backed by Amazon Elastic Block Store (EBS) but limits storage to 50 GB due to per-user costs, creating constraints for users and the community. (2) Configuring and maintaining compute environments remains a technical challenge. Guidance on defining and running reproducible workflows, for example, binder-ready repositories, is often perceived as technically advanced by hackweek participants. (3) Data-adjacent computing requires co-located datasets, which is often not the case across agencies, and user-contributed data archives, making temporary data transfer, storage, and remote execution necessary even when working in a commercial cloud. (4) Performance limitations occur when accessing data not optimized for S3 (McNally et al., 2026). (5) Growing the CryoCloud

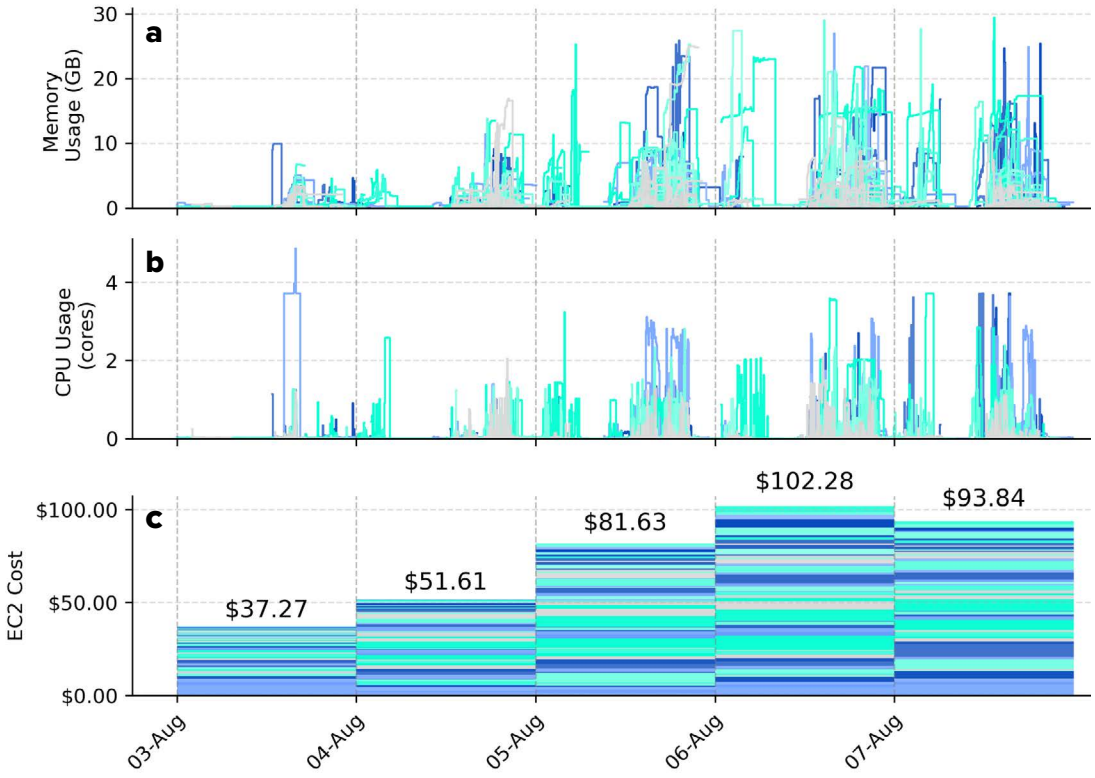

FIGURE 3. Graphs show (a) memory usage (GB) in CryoCloud for each PACE Data Hackweek participant over the five-day 2025 event, (b) central processing unit (CPU) usage (cores) in CryoCloud for each participant, and (c) Elastic Compute Cloud (EC2) cost incurred by each participant over the five-day event. Colors denote individual participants.

community requires additional funding, but clear mechanisms for contributions from grants, institutions, or agencies are not yet established. Future collaboration between hackweek organizers, data engineers, and cloud system administrators would help prioritize and address these challenges, with benefits extending beyond future hackweeks.

While PACE Data Hackweeks have been highly effective in building skills and community, opportunities exist to improve their reach, structure, and long-term impact. (1) Although PACE Data Hackweeks have primarily attracted early-career scientists, they also offer valuable opportunities for late-career researchers to update their skills in modern, cloud-based workflows. Future hackweeks will aim to increase participants from this group to broaden impact across the community. (2) Participant feedback also points to challenges with project development and completion. To address this issue, future PACE Data Hackweeks may adopt a more structured approach, with pre-defined project topics, participant matching, and dedicated mentors to support progress during and after the event.

CONCLUDING REMARKS

Hackweeks have become an effective mechanism for advancing scientific work, in large part because they bring together researchers who might not otherwise collaborate. Interactions often continue after the event, leading to new ideas and follow-on projects. By combining different areas of expertise over an intense but short research sprint, hackweeks provide space to experiment with new approaches and to take risks on novel or uncertain research directions.

PACE Data Hackweek participants not only explored new data products but also developed familiarity with cloud-based workflows and joined a larger community of Earth science researchers working to maximize the benefits of NASA's Earthdata Cloud. Leveraging an established, cloud-based JupyterHub for hackweeks is practical and cost-effective, and provides a nexus for continued interactions. Continued investment in hackweeks and shared cloud resources will be critical for sustaining these benefits, and we plan on hosting PACE Data Hackweeks in the future.

As NASA's Earth Observing System evolves, hackweeks are well positioned to serve as catalysts for skill-building and community formation. The oceanography community faces a mix of opportunities and challenges as it advances open science. Participants in the PACE Data Hackweeks are empowered to translate broad shifts in the data and computational landscape into meaningful, lasting impacts for the community.

SUPPLEMENTARY MATERIALS

The supplementary materials are available online at https://doi.org/10.5670/oceanog.2026.e204.

REFERENCES

Gentemann, C.L., C. Holdgraf, R. Abernathey, D. Crichton, J. Colliander, E.J. Kearns, Y. Panda, and R.P. Signell. 2021. Science storms the cloud. *AGU Advances* 2(2):e2020AV000354, https://doi.org/10.1029/2020AV000354.

Huppenkothen, D., A. Arendt, D.W. Hogg, K. Ram, J.T. VanderPlas, and R. Rokem. 2018. Hack weeks as a model for data science education and collaboration. *Proceedings of the National Academy of Sciences of the United States of America* 115(36):8,872–8,877, https://doi.org/10.1073/pnas.1717196115.

Martin, P.E., E. Eli Holmes, E. Mayorga, J.K. Ansong, U. Bhaskar, J. Cornejo-Donoso, D. Correa-Chilón, R. Damoah, D. Fierro-Arcos, L. Gómez-Navarro, and others. 2025. Harnessing marine open data science for ocean sustainability in Africa, South Asia, and Latin America. *Oceanography* 38(1):59–66, https://doi.org/10.5670/oceanog.2025.121.

McNally, A., L. Sterzinger, and I. Carroll. (2026). Producing Earth science data for impact: Improved commercial cloud usability of archive model data. *Environmental Modelling & Software* 197(106846), https://doi.org/10.1016/j.envsoft.2025.106846.

Mitchell, C., W.-J. Lee, F. Fernandes, J. Gum, A. Kerney, E. Mayorga, T. Moore, N. Mortimer, N. Ribeiro, and V. Staneva. 2026. OceanHackWeek: An inclusive, collaborative approach to developing oceanography data science skills. *Oceanography* 39(1):50–55, https://doi.org/10.5670/oceanog.2026.e104.

Robinson, E., and J. Stewart Lowndes. 2022. The Openscapes Flywheel: A framework for managers to facilitate and scale inclusive open science practices. *Earth ArXiv* preprint, https://doi.org/10.31223/x5cq02.

Scheick, J., A. Barrett, Z. Fair, Z. Katz, L. Lopez, A. Neeley, C. Roberts, B. Smith, T. Snow, R. Wegener, and A. Arendt. 2026. Ten simple rules to increase the usability of Earth observation mission datasets. *ESS Open Archive* preprint, https://doi.org/10.22541/essoar.176892205.56393151/v1.

Snow, T., J. Millstein, J. Scheick, W. Sauthoff, W.J. Leong, J. Colliander, F. Pérez, J. Munroe, D. Felikson, T. Sutterley, and M. Siegfried. 2023. "CryoCloud JupyterBook." Zenodo, https://doi.org/10.5281/ZENODO.7576601.

ACKNOWLEDGMENTS

We gratefully acknowledge the PACE Data Hackweek mentors, tutorial contributors, and science presenters. We thank Tasha Snow for guidance and support with CryoCloud, Eli Holmes for advice on hackweek organization, and Jenny Wong (2i2c) for assistance in analyzing CryoCloud costs and usage statistics. We also acknowledge the ICESat-2 Hackweek team for early scoping support, as well as the OCB Project Office, NASA Ocean Biology Biogeochemistry Program Managers, and the UMBC staff for funding, logistical, and technical assistance. We thank Anthony Arendt and Catherine Mitchell for their thoughtful and constructive reviews of this manuscript. And lastly, we thank all hackweek participants for their enthusiastic engagement.

FUNDING SUPPORT

The PACE Data Hackweek was established with the support of the Ocean Carbon & Biogeochemistry Project Office, which receives funding from the National Aeronautics & Space Administration (NASA 80NSSC25K7397) and the National Science Foundation (NSF OCE-2445578).

AUTHORS

Anna E. Windle (anna.windledipaola@nasa.gov), Ocean Ecology Lab, NASA Goddard Space Flight Center, and Science Systems and Applications Inc., Greenbelt, MD, USA. **Ian Carroll**, Ocean Ecology Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA, and GESTAR II, University of Maryland Baltimore County, Baltimore, MD, USA. **Carina Poulin** and **Skye Caplan**, Ocean Ecology Lab, NASA Goddard Space Flight Center, and Science Systems and Applications Inc., Greenbelt, MD, USA. **P. Jeremy Werdell**, Ocean Ecology Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA.

ARTICLE CITATION

Windle, A.E., I. Carroll, C. Poulin, S. Caplan, and P.J. Werdell. 2026. The PACE Data Hackweek launches a new wave of open science: How we did it (twice), and why. *Oceanography* 39(2):62–67, https://doi.org/10.5670/oceanog.2026.e204.

COPYRIGHT & USAGE

This is an open access article made available under the terms of the Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format as long as users cite the materials appropriately, provide a link to the Creative Commons license, and indicate the changes that were made to the original content.