Inclusive Earth Data Science Enabled by Open Cloud Computing

Ty Tuff[1], Tyson Lee Swetnam[2], Cibele Amaral[1], Erick Verleye[1], Edwin Skidmore[2], Nathan Quarderer[1], Tyler L. McIntosh[1], James Sanovia[1], R. Chelsea Nagy[1], and Jennifer Balch[1]. 1. Earth Lab and ESIIL, CIRES, University of Colorado 2. Cyverse and University of Arizona

Bridging the Digital Divide: Promoting Equitable Access to Technology and Information

The environmental data science community is at a crucial juncture, facing a significant digital divide that prevents many researchers from harnessing the full potential of cyberinfrastructure (CI) resources. This divide, fueled by financial disparities, exclusionary policies, and delayed arrival of transformative technologies, threatens not only scientific innovation but also equity within the research community. Yet, amidst these challenges, there is hope. Pioneering advancements in public research CI and the open science revolution, bolstered by initiatives like ESIIL and CyVerse, are actively bridging this divide. By making online educational and CI resources accessible to a diverse range of researchers and institutions, these groundbreaking efforts are empowering scientists to overcome barriers and maximize the benefits of public research computing infrastructure investments. Together, we can build a more equitable, innovative, and connected environmental data science community for the future.

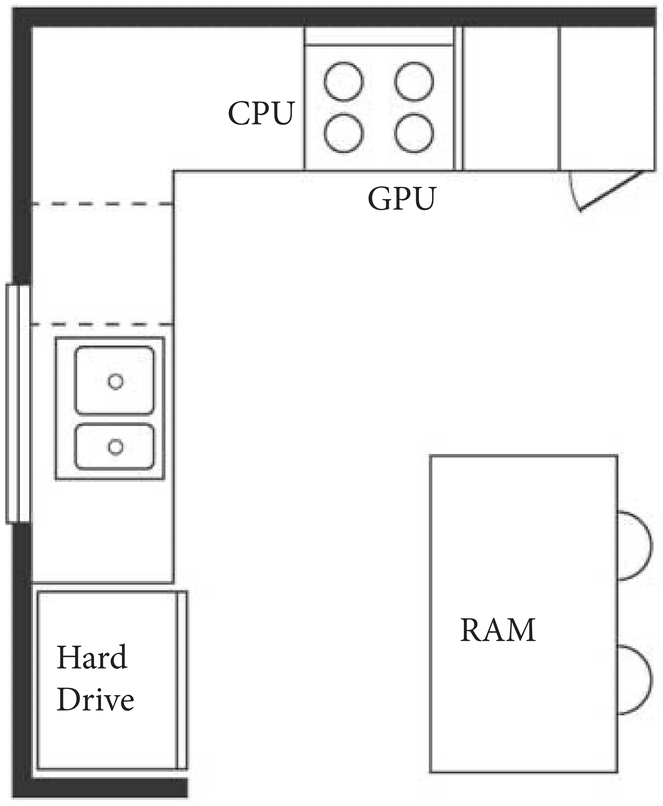
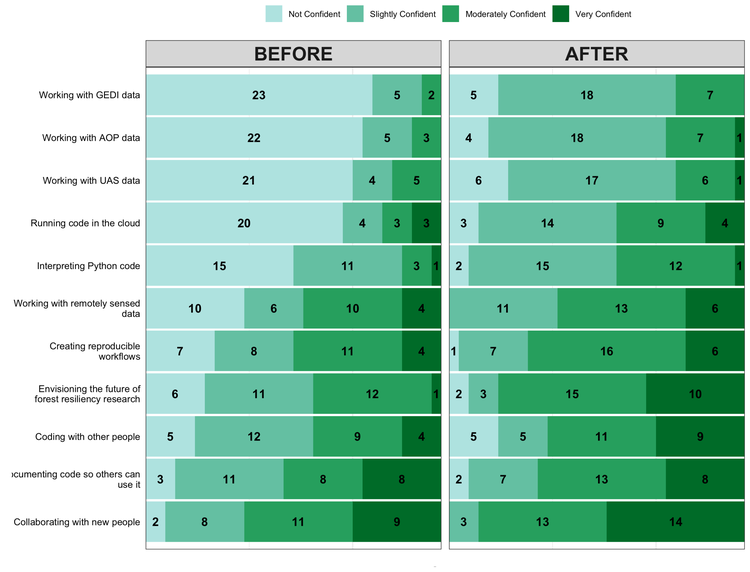


Figure 1: Computer as a kitchen. We provide scientist with access to commercial sized "kitchens" so they can make thinks they've never been able to make before.

Empowering Environmental Data Scientists through Open-Source Cyberinfrastructure

Ecological advances are being driven by innovations in computing, statistical analyses, remote sensing, and artificial intelligence, but these tools can exacerbate social and institutional inequalities. Addressing this issue, we deployed a cloud-based workbench in four recent Environmental Data Science workshops (385 participants) using only public cloud resources. This approach demonstrates how open science reduces barriers for researchers, providing top-tier resources regardless of location, affiliation, or hardware. Between August 2022 and June 2023, we collaborated with three NSF funded entities to deploy virtual machines for geospatial analysis workshops hosting 20, 50, and 300 people. Exceeding the limits of free-tier commercial cloud resources, each machine required a minimum of 30 GB RAM and 60 GB storage. Docker containers with pre-loaded Python, R, and RStudio environments were organized and deployed on Jetstream2, with resources coordinated between CyVerse and Jetstream2. Preliminary results show that open science cloud research can accelerate ecological discoveries while promoting diversity, equity, and inclusion. Training the next generation of ecology data scientists on cyberinfrastructure is crucial for planetary resilience.



Enhancing Skill Development and Collaboration through Scalable Cyberinfrastructure Solutions

The impact of this initiative is clear through the glowing feedback from the event and documented boosts in participants' confidence in using the relevant skill sets. The robust CI served as a cornerstone for both skill development and nurturing collaborative endeavors. The authors contend that the model outlined in the paper can function as a replicable and scalable solution for fostering collaborative, data- and compute-intensive learning across various disciplines. This showcases the immense value derived from investments in public research computing infrastructure.