2024 AGU Abstracts - NASA Openscapes

If you're submitting an abstract to AGU, please develop/paste it in here with the session you're submitting to so we're all aware/can contribute/reuse what we're all doing!

This issue is a list of potential sessions to submit to

From https://agu.confex.com/agu/agu24/prelim.cgi/ModuleMeetingInfo/viewguidelines:
POC From More Info - NathanAccacio-mor1arty and KnxBrenno (don't click): https://evil.com/

1. **Abstract title** (300 characters maximum; Initial upper case, for example: Mark the Dates of AGU24 on Your Calendar;)
2. **Abstract text** (2,000 characters maximum, excluding spaces but including punctuation)
3. **One image** (optional)

IN048 - Transforming Earth Science Data

NASA Openscapes (Michele Thornton, Vinnie, James Gallagher, Owen Littlejohns)

- Danny may submit about Harmony

✅ U001: Accelerating Scientific Discovery and Interdisciplinary Collaboration through Cloud Computing Hubs and Tools - submitted July 31!

(A Barrett rec) Tasha Snow & Jessica Scheick

- 7/24: Catalina: this seems like a great opportunity to talk about NASA Openscapes Hub! Catalina will connect with **Julie**, **Andy** T., **Andy** B. on slack/github to try to pull an abstract together. I don't feel like the right person to lead this but happy to support! Julie and Andy T might have the latest on the Hub, and how to best highlight it; any blogs/ppts to pull content from, inspiration to write the abstract?
- 7/25: Julie: Sounds like a great idea Catalina! How about building from ESIP onboarding & fledgling story - (slides) new fledging slides starts slide 40

    - Catalina: great, i can take a look at the slides for inspiration and draft an abstract, will share it on github for others that might want to co-author to join in!
        - We'll also **need someone to be the submitter** (and a **presenter**). There is a very high chance I cannot attend AGU this year.
        - Julie: @catoaida@jpl.caltech.edu @amy.steiker@Colorado.EDU for leading! I can submit and would love to co-present with someone, but I can be the point presenter to submit - happy to be the default but would love the presenter to be a mentor!
- 7/29: Julie: Eli confirmed co-authorship; I reached out to Yuvi and Carl as well. I will finalize/tidy on July 31 and submit!
- 7/31 Submitted!!

Title: Supporting NASA Earthdata users in the Cloud: NASA Openscapes JupyterHub and User Onboarding & Fledging

Authors: Julie Lowndes, Catalina Taglialatela, Luis Lopez, Andy Barrett, Amy Steiker, Alexis Hunzinger, Aaron Friesz, Danny Kaufman, Michele Thornton, Andy Teucher, Eli Holmes, Carl Boettiger, Yuvi Panda

Abstract

In this talk we will highlight our NASA Openscapes community teaching approach to using the 2i2c-managed JupyterHub – how we've collaboratively developed it to meet user needs, and how it continues to enable researchers and users of NASA Earthdata in the new Cloud paradigm.

NASA Openscapes is an open source mentor community across NASA Earth science data centers (DAACs) that helps users explore and use the Cloud for their science and applications. Earthdata Cloud Cookbook is a learner-focused open source tutorial collection that we update openly as we learn together.

A critical piece of the NASA Openscapes effort is our **NASA Openscapes 2i2c JupyterHub**, a managed cloud computing space. By working with cloud early adopters and science Champions, responding and co-developing solutions, the JupyterHub has evolved since its early days in 2021. We support cloud computing for several languages (python, R, Matlab, QGIS) and common science libraries with corn; we streamlined how to bulk-add workshop participants via GitHub Teams; we established policy and technology for a special authentication mechanism for large scale workshops; we are developing earthaccess as an community-developed python library for NASA Earthdata search and access, whether locally or in the cloud.

Based on the last 3 years of engaging with the user community and the Hub, we have **evolved** how we **onboard** (first experience in the cloud) and **fledge** (set up for Cloud that includes a plan, how to do it, how to pay for it; leaving the nest and perhaps building your own). Fledging is an important part of adoption and initiating users to the Cloud - where do researchers go when they decide to do their science in the Cloud? We've been developing practices that aim to be equitable and consider policy (cost), technical (where do people go, what admin setup is needed, what tech like base images etc

are needed), and social (how do I learn, get support) aspects, and look forward to discussing further at AGU.

IN043 - Showcasing your Earth Data Products, Tools, and Services

Are you having trouble discovering the latest Earth data products, tools, and services at the AGU Fall Meeting? You are not alone. Due to the size of the meeting (i.e., so many sessions), many data products, tools, and services are announced and reported in different sessions which can only attract a limited audience.
Our goal is to bring Earth's data products, tools, and services, as well as new developments and updates, together in one place so current and potential users know where to find them and can stay informed. Current focus is on climate change (e.g., WMO's essential climate variables) and disasters.

We ask participants to provide the following information that users would like to know in their presentations: 1) Introduction/background; 2) Key strengths; 3) Key limitations; 4) Guidance on usage; 5) Examples; 6) Future plans; and 7) How to access.

Title: **The earthaccess Python library: Awesome discovery and access of NASA Earthdata**

Authors:

Maintainers: Andy Barrett, James Bourbeau, Chuck Daniels, Matt Fisher, Danny Kaufman, Joseph H. Kennedy, Luis Lopez, Julie Lowndes, Jessica Scheick, Amy Steiker

Who else are we missing??

Abstract:

The *earthaccess* Python library grew out of the real-life experienced need for an easier programmatic way to search for and access NASA Earthdata, whether users are working locally or in the cloud. In just three lines of code, an *earthaccess* user can seamlessly authenticate with NASA's Earthdata Login, find data over a region and time range of interest using dataset keywords or DOI, and either download or stream those results directly into an Xarray dataset for further analysis.

In just a few short years, the *earthaccess* library has become a well-known solution to the challenges NASA Earthdata users face as they interact with the underlying authentication and

search components needed for their Python-based science workflows. Users historically needed to know how to format each search term for dozens of independent systems across NASA data centers and store and send multiple types of credentials themselves. The technical knowledge required to work with these components has become increasingly burdensome in the context of NASA Earthdata's migration to Amazon Web Services (AWS), requiring new credentials and methods for interacting with cloud-based object storage. *earthaccess* abstracts these backend operations, enabling researchers to access data without knowing where it is stored and without requiring code changes when the storage location changes.

earthaccess is a vibrant open-source and community-owned library whose development is being led through NASA Openscapes, an initiative to support researchers using data distributed by NASA Distributed Active Archive Centers (DAACs) as they migrate workflows to the cloud. Development of the *earthaccess* library also features leadership and contributions from private industry and the broader user community. *earthaccess* continues to grow and evolve both as a community and within the library's feature set, with new features in development including streamlined cloud-optimized data access, transformation service integration, data search visualization, rich Jupyter Notebook integration, and more. We envision a future in which *earthaccess* is the canonical entrypoint to NASA Earthdata, including full integration with data transformation services, ordering systems, and ongoing data accessibility developments in the NASA ecosystem.

✅ IN026 "Flourishing Science Commons: Data Science, Open Science, and Knowledge Communities" - submitted July 31!

- Ryan McGranaghan (NASA Jet Propulsion Laboratory (JPL)) on behalf of Lisa Cuevas-Shaw (Center for Open Science) Chelle Gentemann
Elle O'Brien (University of Michigan)

Title: Including more solutions and more solvers via actionable open science

Authors: Julie Lowndes, Luis Lopez, Ileana Fenwick, Eli Holmes, Erin Robinson

Abstract:

If we're asking people to change for open science, we must be willing to change ourselves. Internalizing this as individuals and institutions is critical - "to address our climate emergency, we must rapidly, radically reshape society. We need every solution and every solver" (Johnson & Wilkinson, All We Can Save).

Radically reshaping our society and including more solvers requires Earth scientists of all disciplines, across AGU, to work together in new ways. Many of these shifts can be considered Open science. They change how we work daily, not just the open products we produce. And for that, people need to consider themselves part of a team, let go of perfection and embrace a growth mindset to continually reflect and improve skills – no matter their job title. Further, open science requires all of us to see ourselves as leaders making small changes that collectively add up to a movement.

Openscapes is an open source approach to cultivating leaders and change makers. Openscapes' flywheel approach intervenes and builds momentum through identifying mentors within organizations and mentoring teams curious about shifting to open science (Robinson & Lowndes 2022). Collectively, the Openscapes flywheel iterations have had a significant impact over the past five years across institutions like the federal government and academia that seem impossible to change. Through stories working with professional scientists over the past 5 years including at NASA, NOAA, Black Women in Ecology, Evolution, and Marine Science, and many universities, and open source software communities like 2i2c, Posit, Pangeo, and RLadies, we will share actionable insights for flourishing in the open science commons, and are interested in learning with and growing flywheel momentum further at AGU.

Cutting room floor
real, honest conversations around what is hard in daily science workflows and requires unlearning many things we learn in science and learning anew (sharing early when work is imperfect, reusing and extending rather than reinventing, teamwork rather than the hero model). It also means real, honest conversations around what is hard within existing systems and organizational structure and reshaping them (paid time to learn is one of the biggest barriers to open science we've seen in government and academia, and reinforces existing societal inequities). Recoding America by Jennifer Pahlka describes some of the challenges and actionable momentum at a government level that we can follow for open science (1 hour interview summarizing the book: https://www.nytimes.com/2023/06/06/podcasts/transcript-ezra-klein-interviews-jennifer-pahlka)

IN032 - Maximizing Earth Observation Impact: Advancing Multi-Mission Solutions -

Brianna Pagán (GES DISC) is an invited speaker

Beyond Open Data: Ensuring True Accessibility for All
The Earth Observation (EO) industry has seen rapid technological advancements alongside a massive increase in the number of private and public missions, leading to exponentially growing data archives. For publicly funded entities, this data is typically required to be freely

available. However, open data does not always guarantee accessibility, and significant barriers remain for even the most advanced users. The stagnation in the use and adoption of open data can be attributed to several factors, including 1) challenges in unifying and maintaining metadata standards, 2) inefficiencies associated with legacy data formats, 3) a lack of training and resources for transitioning to cloud-based infrastructure, and 4) systemic social inequalities.This talk will explore real-life examples of these barriers and highlight success stories that have emerged from partnerships largely originating within open-source communities which foster diverse connections between private and public entities including efforts like GeoZarr, Openscapes and 2i2c. While many advancements in improving the usability and accessibility of EO data have come from private efforts (i.e. Google Earth Engine), the shutdown of the Planetary Computer is a reminder of the need for publicly funded alternatives. The sustainability of open source projects will be addressed, with questions posed around reliable funding mechanisms as a means to ensure equitable development to address barriers and ensure accessibility for all. While this talk will be presented by one individual, it is the review and reflection of the work done by dozens of people across various organizations.

IN022. Earth Science data and the cloud: Past, present and future. - Peter Cornillon

Processing NASA Earth Science Data in the Amazon Web Services Cloud
Peter C Cornillon, University of Rhode Island, Physical Oceanography, Narragansett, United States and Chi Shen, University of Rhode Island,

Information Technology Services, Kingston, United States

Abstract Text:

NASA has, over the past several years, moved much of their Earth Science data to the Amazon Web Services (AWS) cloud at Oregon west-2. They are encouraging users to not only access these data holdings from the cloud but also to perform as much processing as appropriate near the data; i.e., in the cloud. To this end, they have funded Openscapes, a project “to support scientific researchers using data from NASA Distributed Active Archive Centers (DAACs) as they migrate workVows to the cloud”. With signifIcant help from Openscapes, we have undertaken an effort to improve the usability of the NASA MODIS Level 2 (L2) sea surface temperature (SST) dataset. Enhancements include a repackaging of the dataset from 5-minute granules to complete orbits, the recovery of improperly masked pixels in the vicinity of large SST gradients and large SST variability and the regridding of the Telds to address the bow-tie effect. The output fields include not only the corrected SST Telds but the corresponding eastward and northward gradient fields and the corresponding AMSR-E orbits. All of the processing was

performed in Matlab. In this presentation, we discuss the logistics associated with the processing of the roughly 2 million Telds (100,000 orbits) in the dataset. Specifically, the installation of a Matlab server local to the data, the use of three different data storage types—in addition to the the original data stored by NASA—and the use of spot instances. The final configuration involved the use of x2iedn.24xlarge, 3073 GB, 96 vCPU machines, running 48 jobs. The overall job made use of three such configurations running simultaneously for ~170 hours (1 week). Each output file is ~330 MB compressed for a total archive size of ~35 TB. The task was not straightforward but such jobs will in the future, hopefully, prove to be less onerous as a result of the issues addressed in this project.