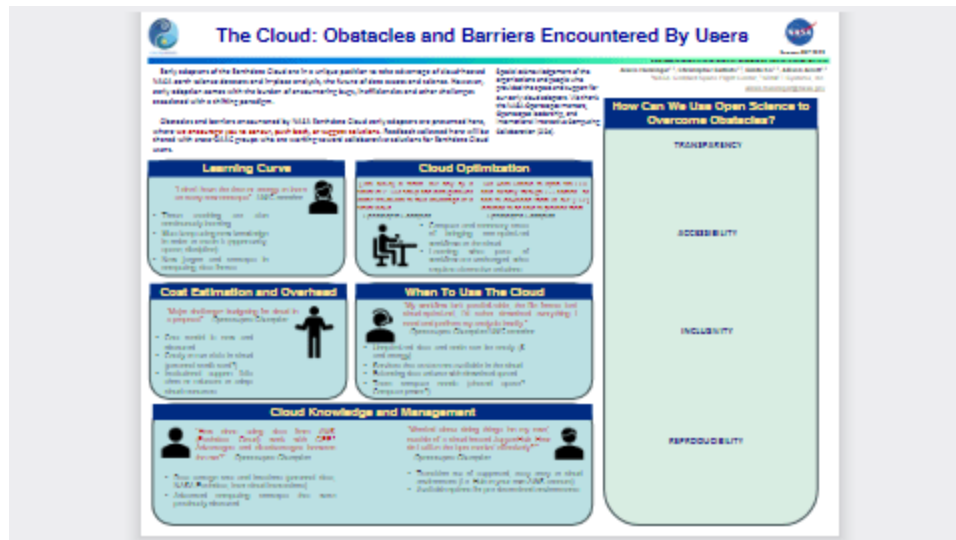


The Cloud: Obstacles and Barriers Encountered by Users



Alexis Hunzinger, Christopher Battisto, Allison Alcott, Binita KC

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC)



PRESENTED AT:



WRECK THIS POSTER!

Early adopters of the Earthdata Cloud are in a unique position to take advantage of **cloud-hosted** NASA earth science datasets and **in-place analysis**, the future of data access and science. However, early adoption comes with the burden of encountering inefficiencies and challenges associated with a shifting paradigm.

Obstacles and barriers encountered by NASA Earthdata Cloud early adopters are presented here, framed by direct quotes from users, and ***we encourage you to concur, push back, or suggest solutions*** using the [Comment](#) feature or reaching out to the corresponding author.

Feedback collected here will be shared with other groups who are working toward collaborative solutions for Earthdata Cloud users.

"I DON'T HAVE THE TIME OR ENERGY TO LEARN SO MANY NEW CONCEPTS."

Learning something new requires **opportunity, space, time, discipline** and **repetition**.

Many data users cannot allot the time and energy required to get up to speed on new concepts and terminology related to cloud storage & computing.

Those *teaching* cloud concepts are also continuously learning, making the learning process for a user iterative - there is always something new to learn!

How others overcome this obstacle

Sharing struggle and success stories to learn from others' experience

Kind teaching tip: Avoid assuming knowledge ("just", "easy") or compute system

"MAJOR CHALLENGE: BUDGETING FOR CLOUD IN A PROPOSAL."

Cost is the primary concern we hear from users introduced to the cloud. For them, the cost model is new and obscured, making budgeting difficult.

Trialing workflows in the cloud causes angst, because **even "free trials" require a credit card** upon registration and allotted computing and storage may not be adequate for testing.

Some users feel their **institutions are reluctant** to adopt the cloud, so they must lean on personal resources to experiment.

Even if a user's institution *does* fund cloud computing, there is a transition period where both existing compute and storage and cloud compute and storage are being used. This **divides attention and funds** between both platforms.

Differing opinions from users: does cloud storage and computing increase equity?

Yes! Cloud is a solution for equity, as it provides under-resourced organizations (e.g. small universities, international groups) access to equipment they don't own or maintain.

vs.

No! Larger organizations, who already own and maintain equipment for computing, often have the budget to also use cloud storage and computing. This increases their pool of resources, giving them a larger advantage.

How others overcome this obstacle

AWS Free Tier offerings (<https://aws.amazon.com/free/>). AWS Pricing Calculator (<https://calculator.aws/>)

Using cost managers (<https://docs.aws.amazon.com/cost-management/latest/userguide/what-is-costmanagement.html>) and alerts

Learning from others' experiences (<https://openscapes.org/blog/2023-11-07-coiled-openscapes/>)

"MY WORKFLOW ISN'T PARALLELIZABLE, THE FILE FORMAT ISN'T CLOUD-OPTIMIZED, I'D RATHER DOWNLOAD EVERYTHING I NEED AND PERFORM MY ANALYSIS LOCALLY."

"[On the preference to perform analysis locally] ...this is fine if data wasn't getting bigger!"

These testimonies illustrate a user's need to **consider if and when cloud computing is appropriate** for their workflow. Considerations should include:

- What is the data volume?
- How long will it take to download?
- Can I store all that data?
- Do I have the computing power for processing?
- Does my team need a common computing environment?
- Do I need to share data at each step or just an end product?

"[The cloud] is faster, but only by a factor of 2. Our setup and data granules aren't structured to take advantage of a faster cloud."

Optimizing data and code for usage in the cloud is a challenge for both users that are unfamiliar with the cloud and users who are cloud-savvy.

Users unfamiliar with the cloud quickly meet **new terms and concepts** they must learn and then figure out **how to implement**.

Cloud-savvy users must put in the time and money required to adjust code and datasets to work effectively in the cloud. Sometimes this means that the **time and monetary cost of conversion** exceeds the cost of running the workflow locally.

Q: Who is responsible for cloud infrastructure and data optimization?

Should cloud infrastructure set up and maintenance fall on the scientist or user, or should it be left to experts and institutions? Should a user be responsible for optimizing datasets they do not maintain, or should data stewards offer these?

"WORRIED ABOUT DOING THINGS 'ON MY OWN', OUTSIDE OF A CLOUD-HOSTED JUPYTERHUB."

Users who have the opportunity to trial a cloud environment via a friendly user interface, like a cloud-hosted JupyterHub, find themselves lost after the experience ends.

Prescribed cloud environments are great for learning, but often **obscure important cost and advanced computing concepts**.

Users feel overwhelmed by the many aspects of managing their own resources in the cloud - compute, storage, egress, and cost.

How others overcome this obstacle

Some users attempt to navigate an **AWS account on their own**. Others turn to products and organizations that offer **similar managed resources**.

Examples include:

AWS SageMaker (<https://aws.amazon.com/sagemaker/>) Coiled (<https://www.coiled.io/>) 2i2c (<https://2i2c.org/>)
Digital Ocean (<https://www.digitalocean.com/>)

"HOW DOES USING DATA FROM AWS WORK WITH GEE?"

Questions often arise about using data stored in **different clouds** (AWS, Google Earth Engine, Azure) or even two **different regions** of the same cloud (AWS us-west-2, us-east-1, etc.).

For example, NOAA satellite data are in AWS region us-east-1 and NASA Earthdata are in AWS region us-west-2. Egress, even out of a region, adds complication for users who want to **use data from both agencies efficiently** in the cloud.

TRANSCRIPT

ABSTRACT

Increasing exposure to and adoption of the NASA Earthdata Cloud yields new concerns from users. Obstacles and barriers encountered by users include: financial resource allotment, steep learning curves, institutional support, and cloud-readiness of data.

We present common concerns expressed by early cloud adopters, and encourage conference-goers to relay their own or users' experiences. We also present proposed solutions for some of these challenges.

We encourage brainstorming for how open science principles can help solve some of the barriers and obstacles users encounter with the cloud.

