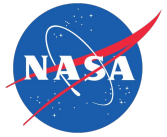


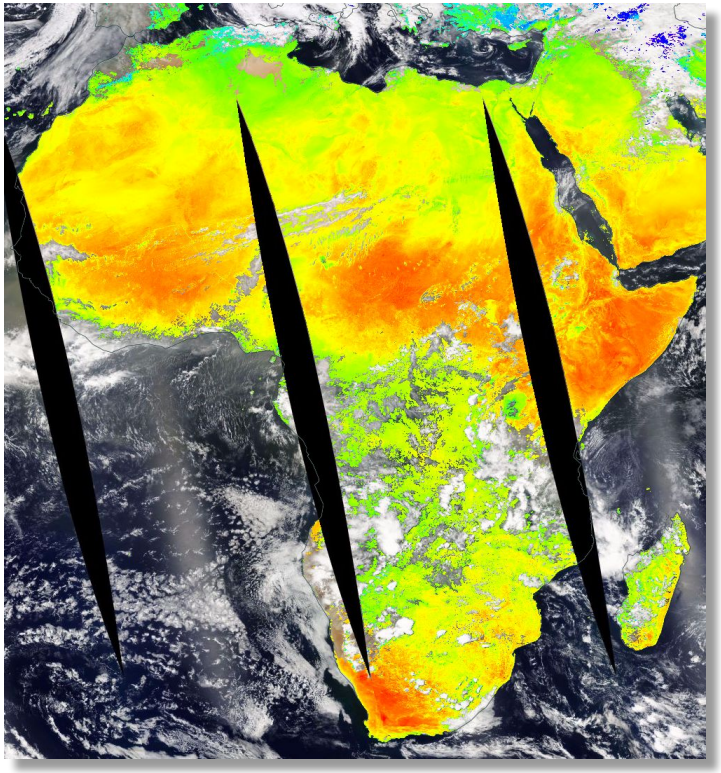
NASA Briefing to Unidata

Christine Smit

Goddard Earth Sciences Data and Information Services Center (GES DISC), Software Engineer

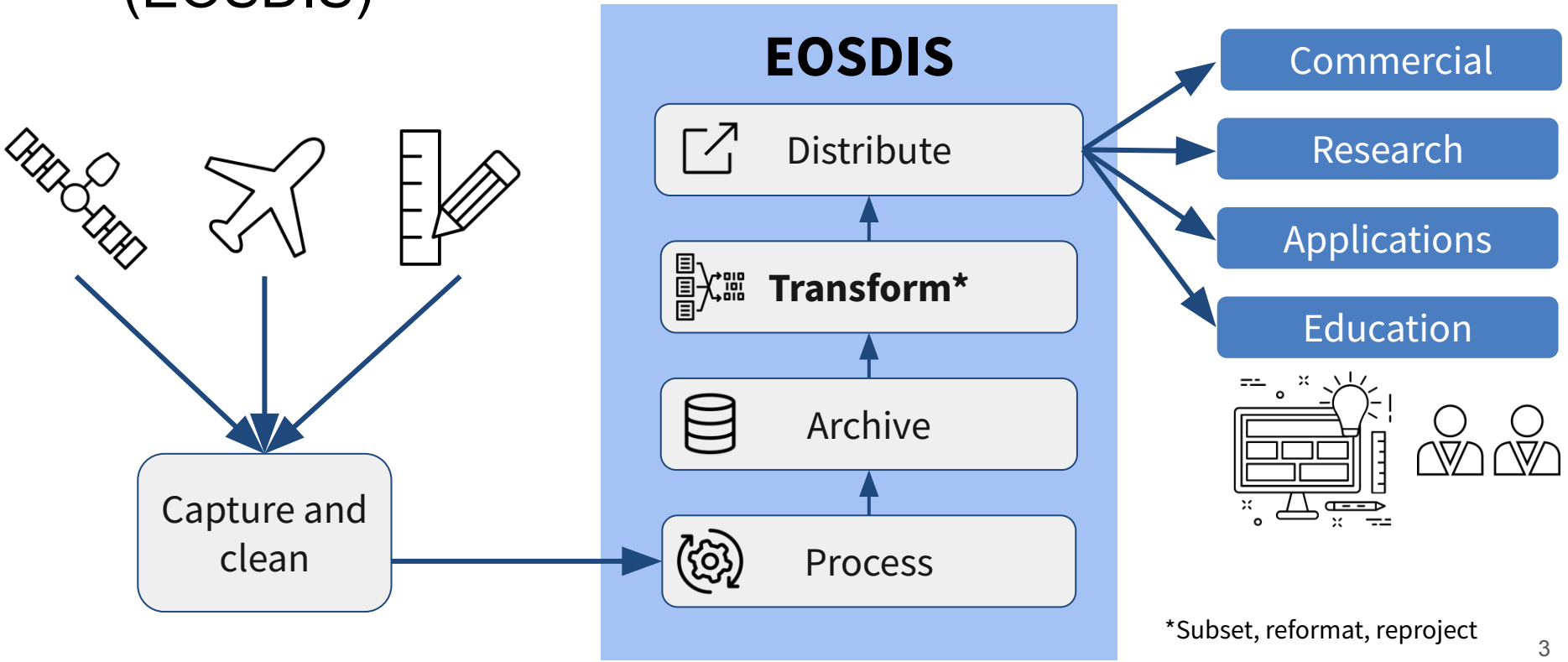
NASA's Earth Science Data Systems Program

- Actively manages NASA's Earth Observations
  - Multivariate
  - Heterogeneous
  - Diverse satellites and suborbital platforms
  - Most spheres: atmo, bio, cryo, hydro, litho
- Designed to support rigorous science **research**
- Processes instrument data to create high quality *long-term* Earth science data records.



http://go.nasa.gov/2mMd5g1

Earth Observing System Data and Information System (EOSDIS)

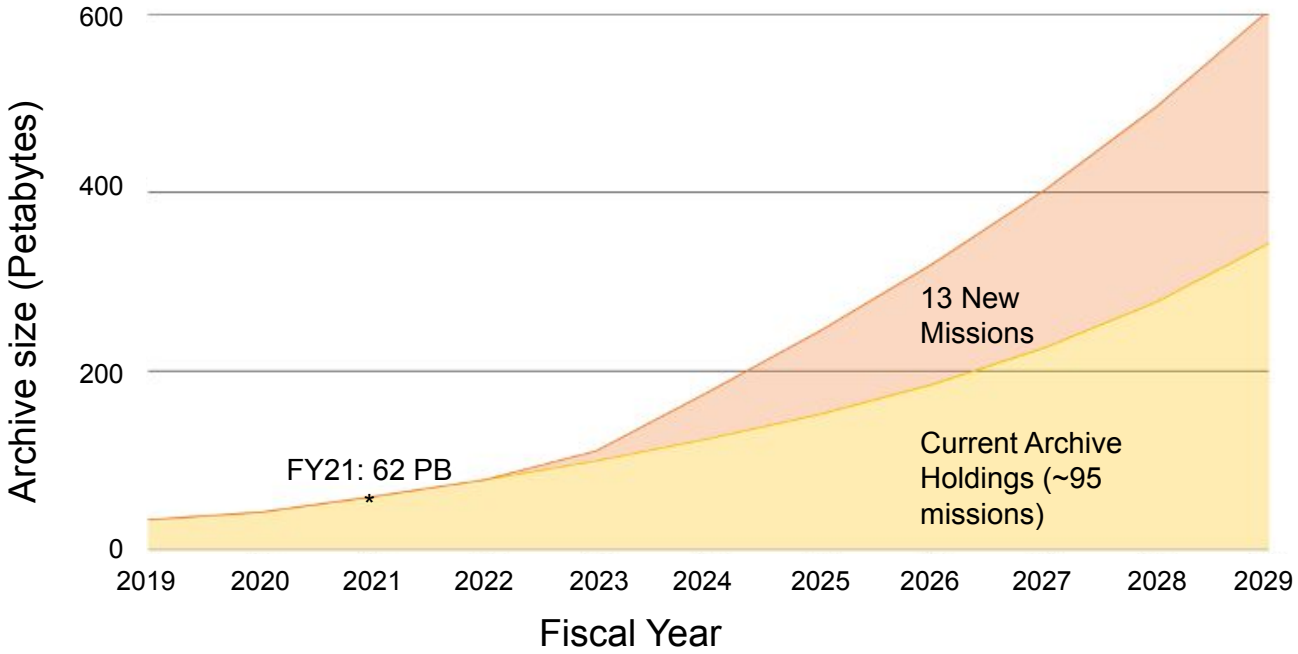


*Subset, reformat, reproject

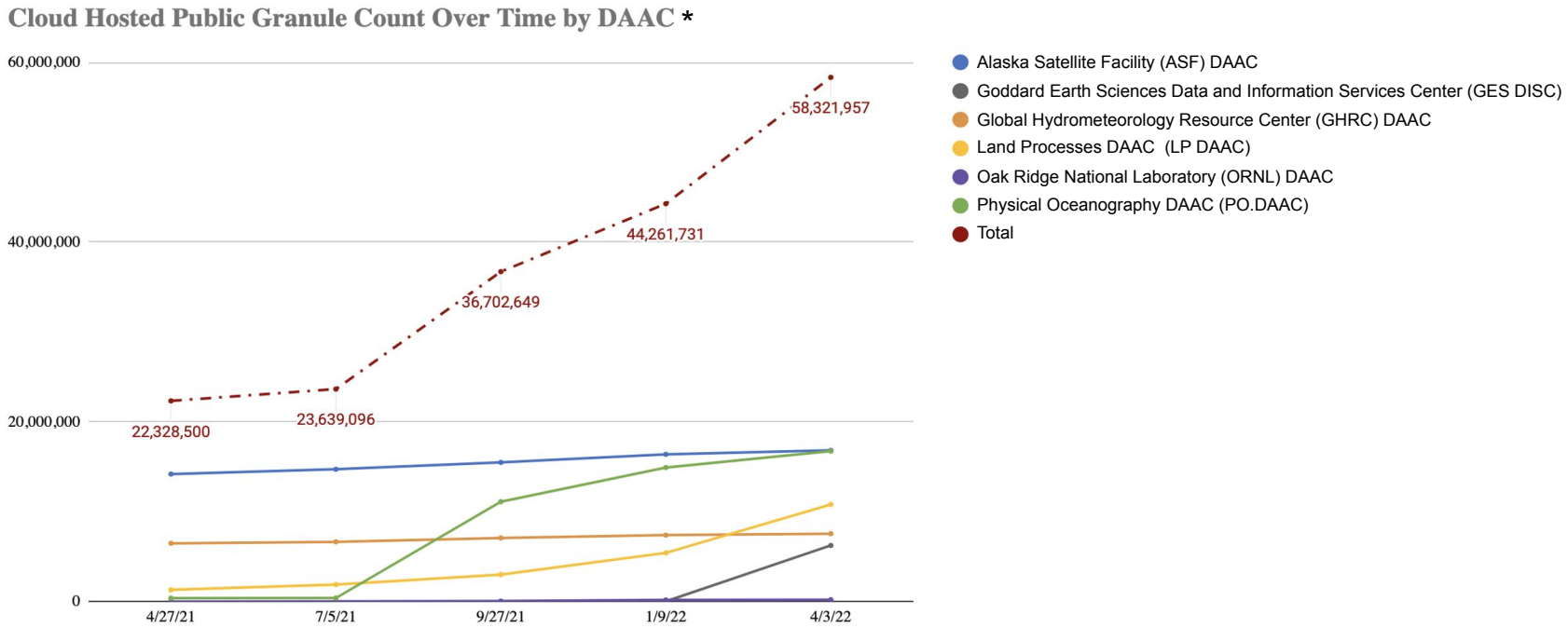
ESDIS in the Cloud: Archive Growth Projection

As the archive grows exponentially, **downloading** entire datasets before working on them becomes **infeasible**.

The best path forward is allowing users to run their **calculations** in the cloud right **next to the data**.



ESDIS in the Cloud: Progress



* Distributed Active Archive Centers

Data gathered from Common Metadata Repository (CMR) inventory on 4/3/22

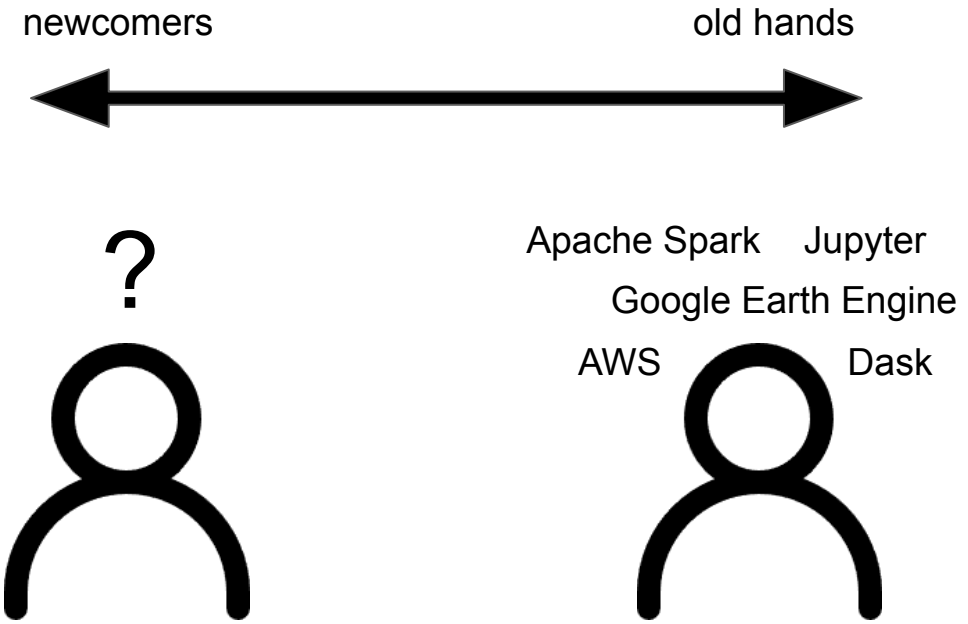
Our users and the Cloud

Cloud newcomers

- Have **heard** about "the Cloud"
- Are **interested** in giving it a try
- Aren't sure where to **start**

Cloud old hands

- Are mostly interested in **where** the data sits and **how** to access it



Closing the knowledge gap for Cloud newcomers

Leap to the cloud involves new

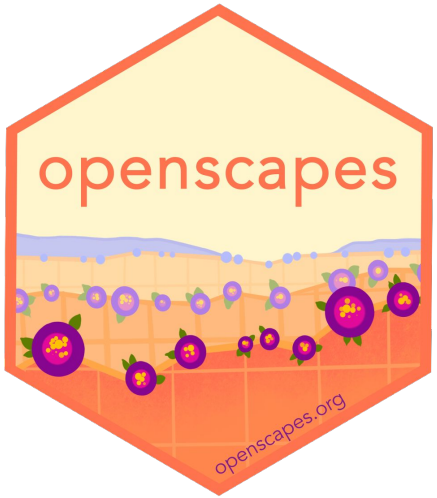
- technologies (jupyter, dask, cloud storage, Elastic Compute Cloud (EC2), lambda, …)
- ways thinking about algorithm development (distributed computing, greater-than-in-memory calculations, network speeds vs. bus speeds, …)
- budgeting frameworks (paying on demand rather than up front)
- security management (networking, credential management, …)

→ It's a larger leap than previous iterations of technology change.

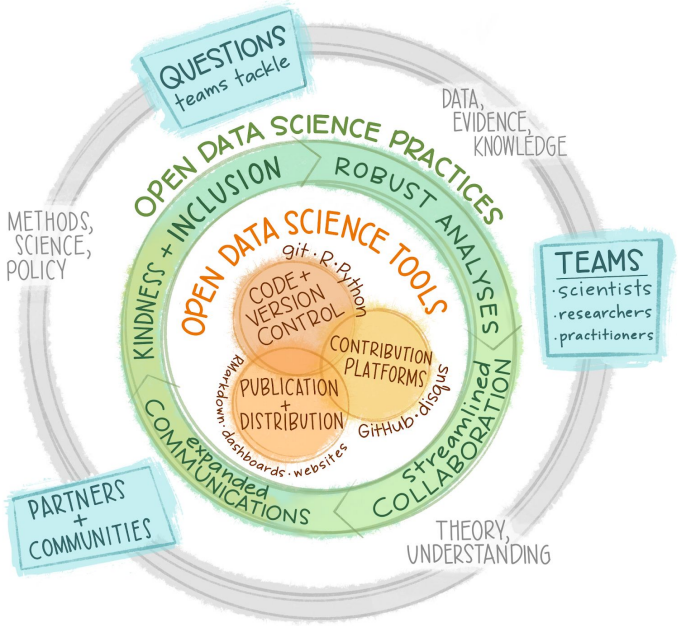
External partner collaboration

Openscapes

- "Believe[s] open practices can accelerate data-driven solutions and increase diversity, equity, inclusion, and belonging in science" (https://www.openscapes.org/approach/)
- Was founded in 2018 by **Dr. Julia Stewart Lowndes** at the National Center for Ecological Analysis and Synthesis with a Mozilla fellowship. Joined by **Erin Robinson** of Metadata Game Changers in late 2020.
- Received NASA Research Opportunities in Space and Earth Science (ROSES) funding in Spring 2021



Openscapes: open data science approach



https://openscapes.org/approach

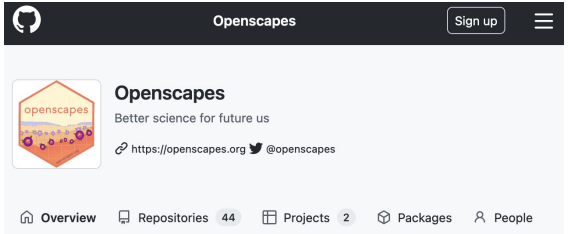
- **Researcher-centered, focused on teams**. Practice and feel safe working openly with yourself and your team; then ease into more.
- **Create space & place to explore & learn.** Cohort Calls, Seaside Chats, Co-Working; GitHub, R, Python, Quarto, Google Drive, Slack; Efficiency Tips & Inclusion Tips.
- **Cultivate relationships & real connections.** Welcoming folks with diverse backgrounds; meeting where they are; skills to empower immediate work; kinder science.
- **Open culture: Learning, teaching, iterating.** Not a checklist - a continual practice. Imperfect, messy. Takes time.

NASA Openscapes

- **Openscapes**
  open data science expertise and infrastructure
- **EOSDIS**
  science data specialists and mentors from Distributed Active Archive Centers (DAACs)
- **Community**
  scientists who use EOSDIS data

https://nasa-openscapes.github.io

NASA Openscapes Technology



https://github.com/Openscapes

- **Source Control: git + github**
  collaboration between DAAC mentors, mentors and scientists, hackathon participants, etc.
- **Compute Platform: Jupyter + python**
  shortened time to working code: no installation required, no compiling
- **Documentation:** Quarto
  beautifully laid out documents with text, images, notebooks, equations, citations, etc.
- **Cloud Access:** 2i2c
  professionally managed JupyterHub in the cloud

NASA Openscapes: first year accomplishments

DAAC mentor development

- 10+ talks & workshops
  - American Geophysical Union (AGU)
  - User Working Groups (UWGs)
  - Surface Water and Ocean Topography (SWOT) Ocean
  - ECOsystem Spacebourne Thermal Radiometer Experiment on Space Station (ECOSTRESS)
  - Train
  - internal DAAC staff
  - Champions cohort
  - …
- Reused tutorials
- Internal and external mentoring sessions

NASA Openscapes: first year accomplishments

Cloud Hackathon: November 2021

- 65 Openscapes 2i2c JupyterHub Amazon Web Services (AWS) instances
- 50 forks of Cloud Hackathon Github repo
- 8 hack-team projects presented on Day 5

March-April 2022: Openscapes Champions Program

- 7 research teams met virtually 5 times over 2 months
- Teams explored open data science and applied their own workflows
- Participants learned from each other: Mentors learn pain points from real science groups and then can invest in software development and teaching