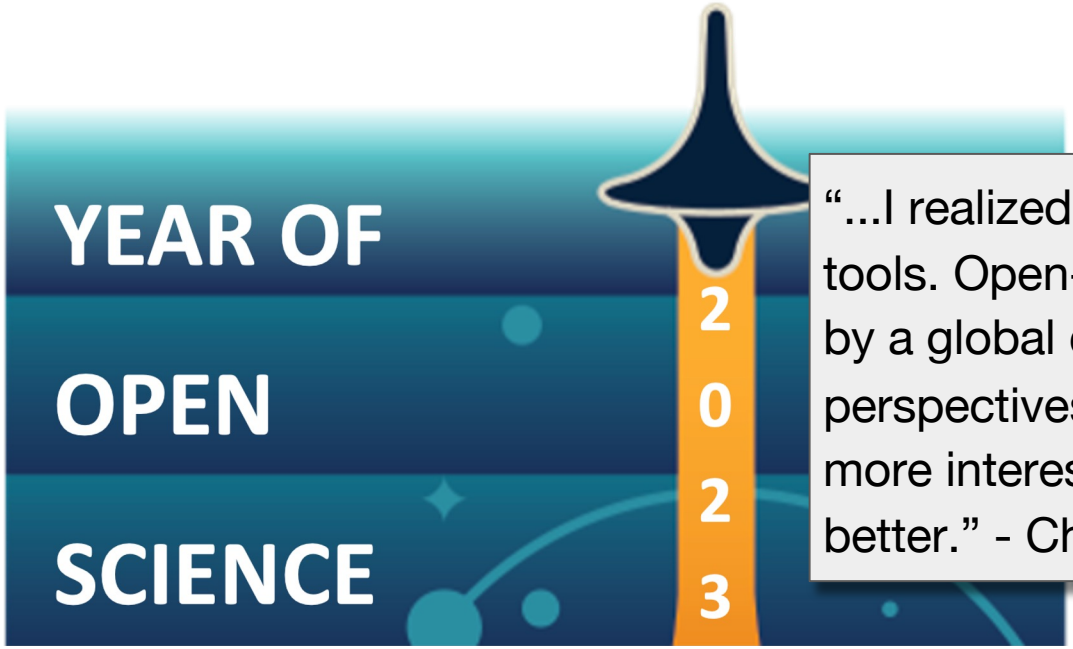
CryoCloud: Accelerating scientific discovery for Cryosphere communities with open cloud infrastructure

Tasha Snow[1], Joanna Millstein[2], Wilson Sauthoff[1,3], Wei Ji Leong[4], James Colliander[5,6], James Munroe[5], Jessica Scheick[7], Denis Felikson[8], Fernando Perez[9], Tyler Sutterley[10], Matthew Siegfried[1,3]

[1] Department of Geophysics, Colorado School of Mines
[2] Massachusetts Institute of Technology - Woods Hole Oceanographic Inst.
[3] Hydrologic Science & Engineering Program, Colorado School of Mines
[4] Byrd Polar Research Center, The Ohio State University
[5] International Interactive Computing Collaboration (2i2c.org)
[6] Department of Mathematics, University of British Columbia
[7] Earth Systems Research Center, University of New Hampshire
[8] NASA Goddard Space Flight Center
[9] Statistics Department, University of California, Berkeley
[10] Applied Physics Laboratory, University of Washington

CryoCloud born within NASA's open science revolution



"...I realized that open science isn't just about tools. Open-science innovation is being driven by a global community with diverse perspectives. The scientific questions are more interesting and nuanced, the solutions better." - Chelle Gentemann

open.science.gov

AGU Advances

RESEARCH ARTICLE
10.1029/2020AV000354

Science Storms the Cloud

C. L. Gentemann[1,2], C. Holdgraf[3,4], R. Abernathey[3,5], D. Crichton[6], J. Colliander[3,7,8], E. J. Kearns[9], Y. Panda[3], and R. P. Signell[10]

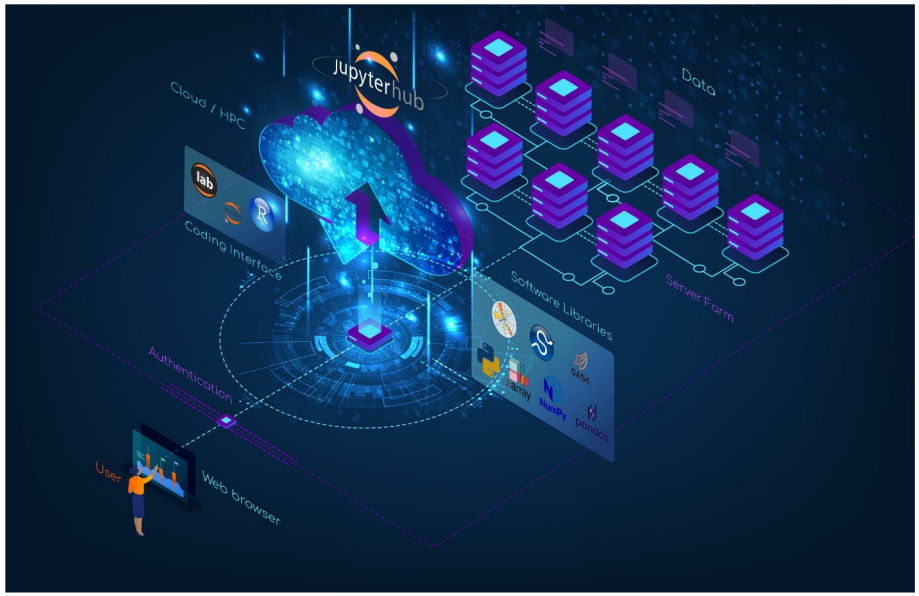
ChelleGentemann

Key Points

Science stands at the cusp of a new, open science, cloud-enabled era

Advances in data, software, and computing are enabling transformational, interdisciplinary science, **changing the realm of possible questions**

Deliberately designed open science communities can **advance science and inclusivity** simultaneously



https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020AV000354

AGU Advances

RESEARCH ARTICLE
10.1029/2020AV000354

Science Storms the Cloud

ChelleGentemann

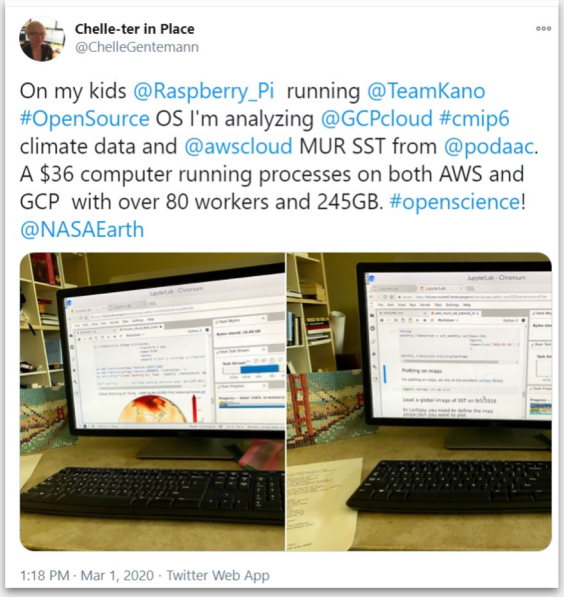
Key Points

Science stands
cloud-enabled e

Advances in dat
transformational
realm of possib

Deliberately des
advance scienc

Process massive amounts of data from a $36 computer…



…or from your cell phone or on a flight

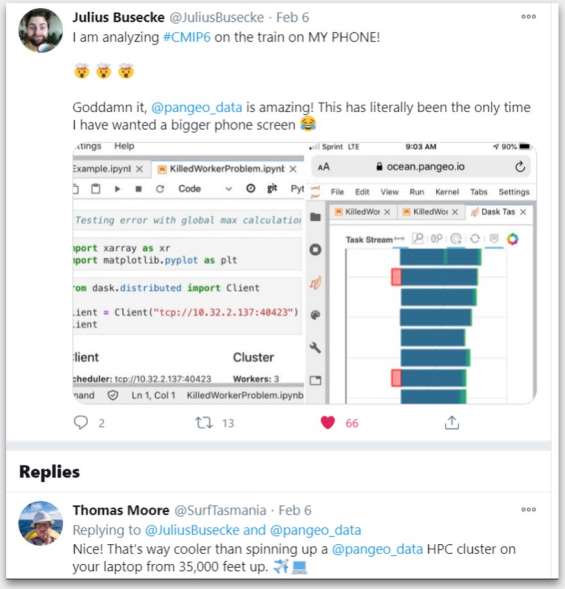


Image credits: Chelle Gentemann

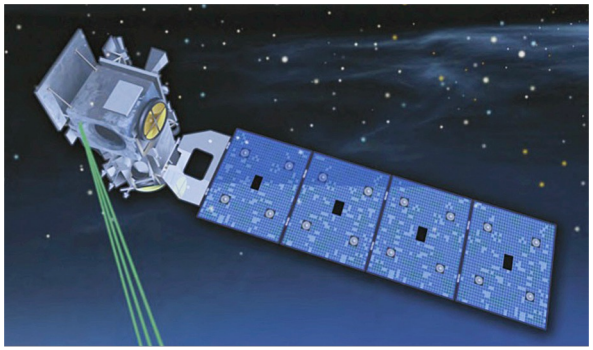
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020AV000354

New spaces and organizational models are needed

Cloud computing and ICESat-2 science

Cloud computing and open science concerns from the May 2022 ICESat-2 Science Team Meeting

- Non-intuitive pricing structures, computing options, infrastructure
- Poor documentation
- Costly to use
- Not obviously more collaborative or faster

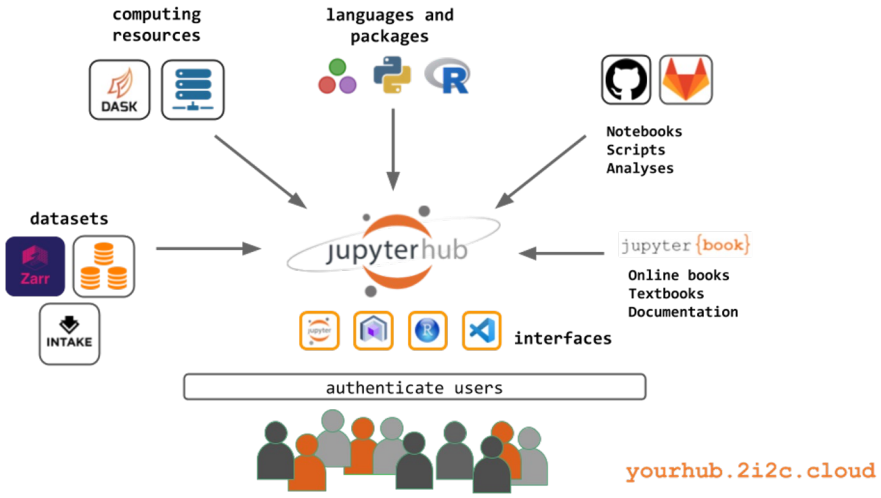


This didn't ring true to our experience in the cloud!

2i2c.org

The International Interactive Computing Collaboration

- **Non-profit.**
- **Service provider** for interactive computing infrastructure.
- An R&D team that **contributes back to open source** communities.



Berkeley Computing, Data Science, and Society

jupyter

binder

PANGEO

Callysto

SYZYGY.CA

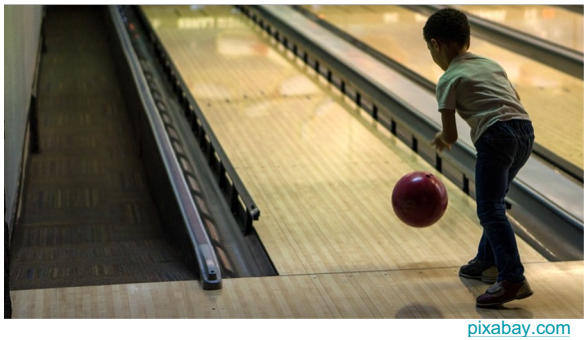
Funding (Open Science Program)

CryoCloud: A cloud-computing platform with *bumpers*

Goal: Simple and cost effective managed cloud environment for training and transitioning new users to cloud workflows and determining community best practices

Built and developed for cryosphere scientists by software professionals at **2i2c** to make it possible to:

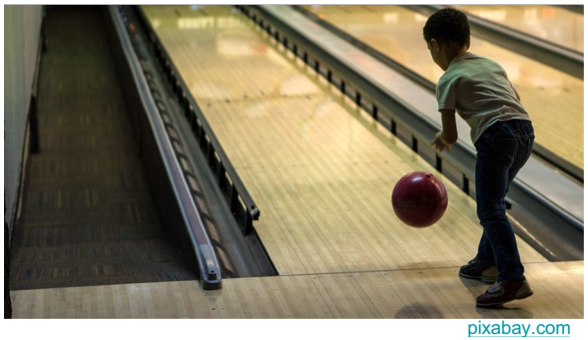
- Process data faster
- Minimize downloading
- Democratize science



pixabay.com

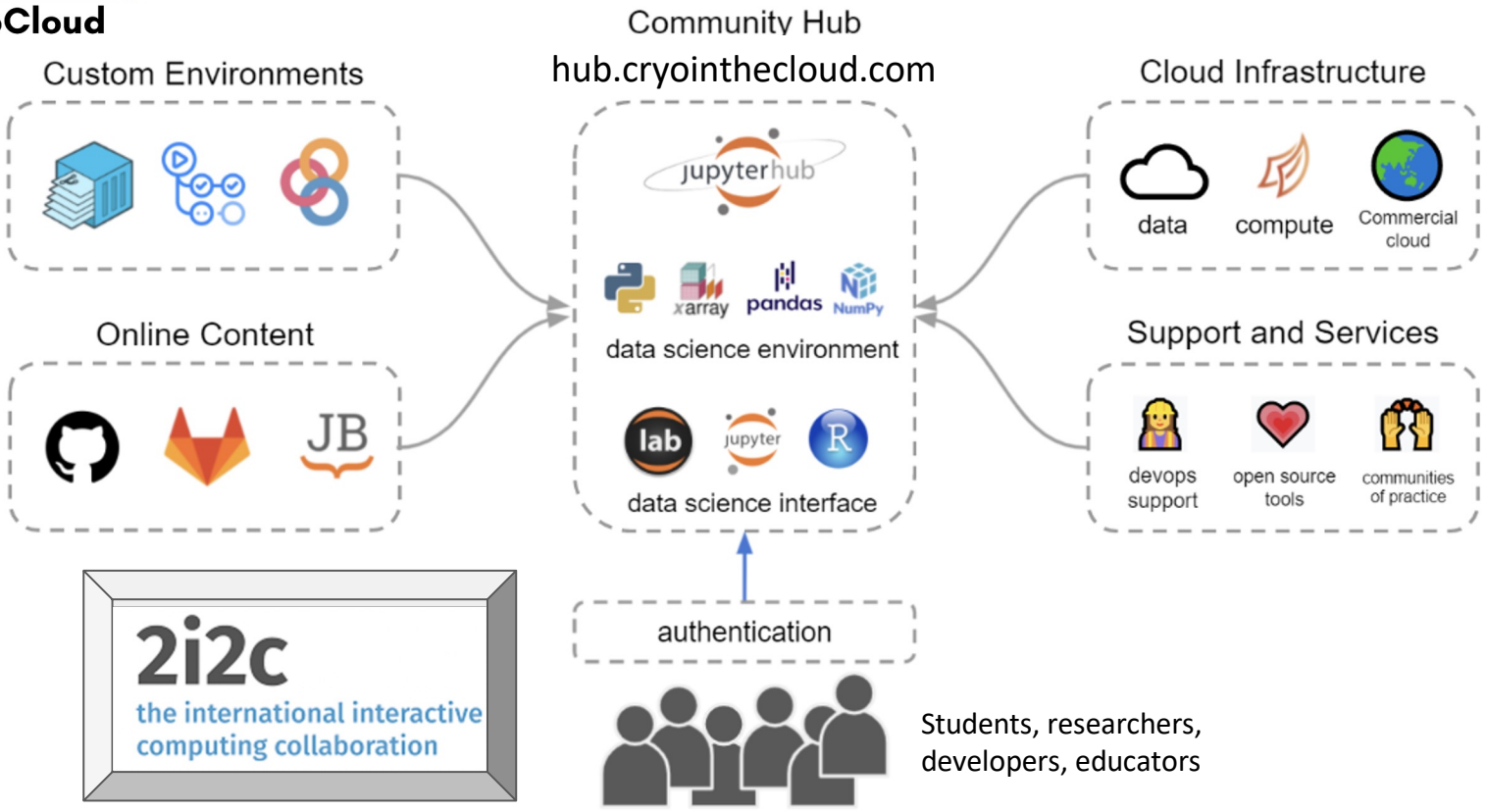
CryoCloud: A cloud-computing platform with *bumpers*

- Persistent for (at least) three years
- Small servers (32 Gb / 4 CPU) for all users, option to bring own cloud credits to access larger servers
- New tool development
  - Personal cost-monitoring tool
  - Intra- and inter-hub collaboration tools
- Helping 2i2c scale with community surveys, feedback, and guidance



pixabay.com

CryoCloud: accelerate discovery and enhance collaboration



CryoCloud community building

cryointhecloud.com

CryoCloud JupyterBook

All the content! More about us, resources, training, and tutorials all found here!

CryoCloud JupyterHub

Get onto the cloud. Our shared cloud platform for NASA Cryosphere communities.

- CryoCloud Github: github.com/cryointhecloud

- CryoCloud Slack
- Community office hours
- Training, tutorials, and resources
- Bring in Cryosphere communities and share in infrastructure ideation and construction

Advantages to a community platform

CryoCloud helps accelerate science and makes open science easy

Data read-ins 1-2 orders of magnitude faster

Easy to use, customizable – same software on your local/HPC/cloud

Collaboration made easy – co-coding, shared tools

Eliminates technology bottlenecks – shared challenges

No software expertise needed – cloud-computing as a service

CryoCloud helps accelerate science and makes open science easy

Data read-ins 1-2 orders of magnitude faster

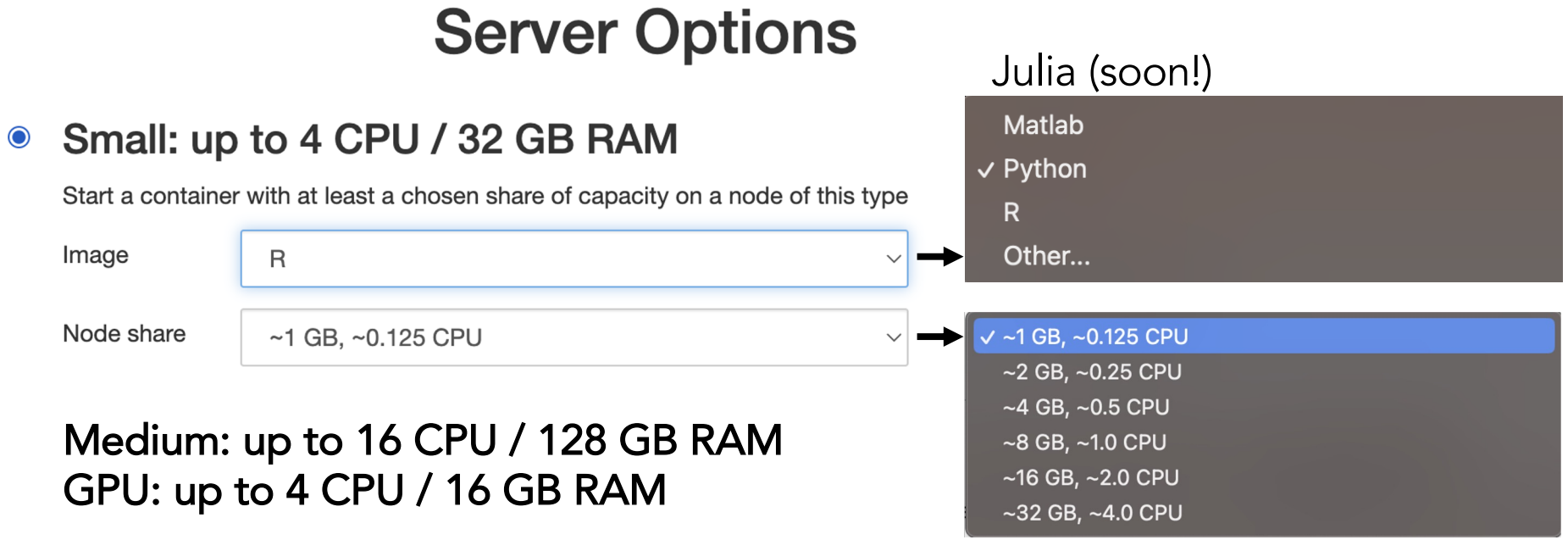
Easy to use, customizable – same software on your local/HPC/cloud

Collaboration made easy – co-coding, shared tools

Eliminates technology bottlenecks – shared challenges

No software expertise needed – cloud-computing as a service

Versatile choice of computer language and server size



Medium: up to 16 CPU / 128 GB RAM
GPU: up to 4 CPU / 16 GB RAM

CryoCloud helps accelerate science and makes open science easy

Data read-ins 1-2 orders of magnitude faster

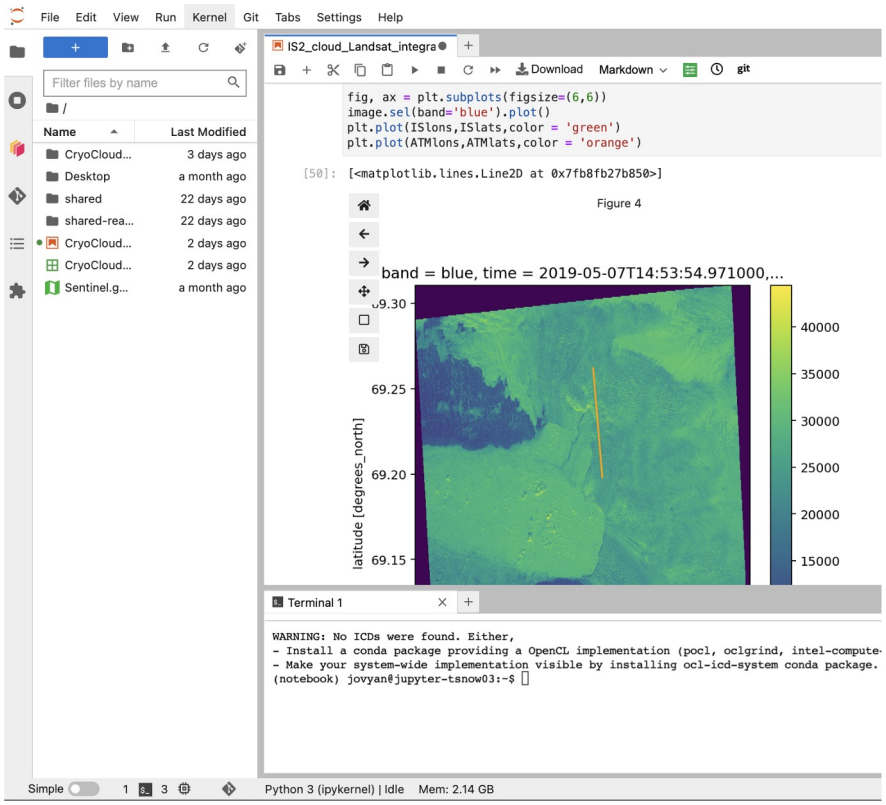
Easy to use, customizable – same software on your local/HPC/cloud

Collaboration made easy – co-coding, shared tools

Eliminates technology bottlenecks – shared challenges

No software expertise needed – cloud-computing as a service

Different kinds of users in one place accelerates feedback and collaboration



Datasets

ICESat-2
ICE, CLOUD, AND LAND ELEVATION SATELLITE-2

NASA MODIS
Moderate Resolution Imaging Spectroradiometer

LANDSAT
LANDSAT-8
USGS • NASA

Tools and Developers

xarray

icepyx

SR

earth access

earthaccess
A Python Library for NASA Earthdata

CryoCloud helps accelerate science and makes open science easy

Data read-ins 1-2 orders of magnitude faster

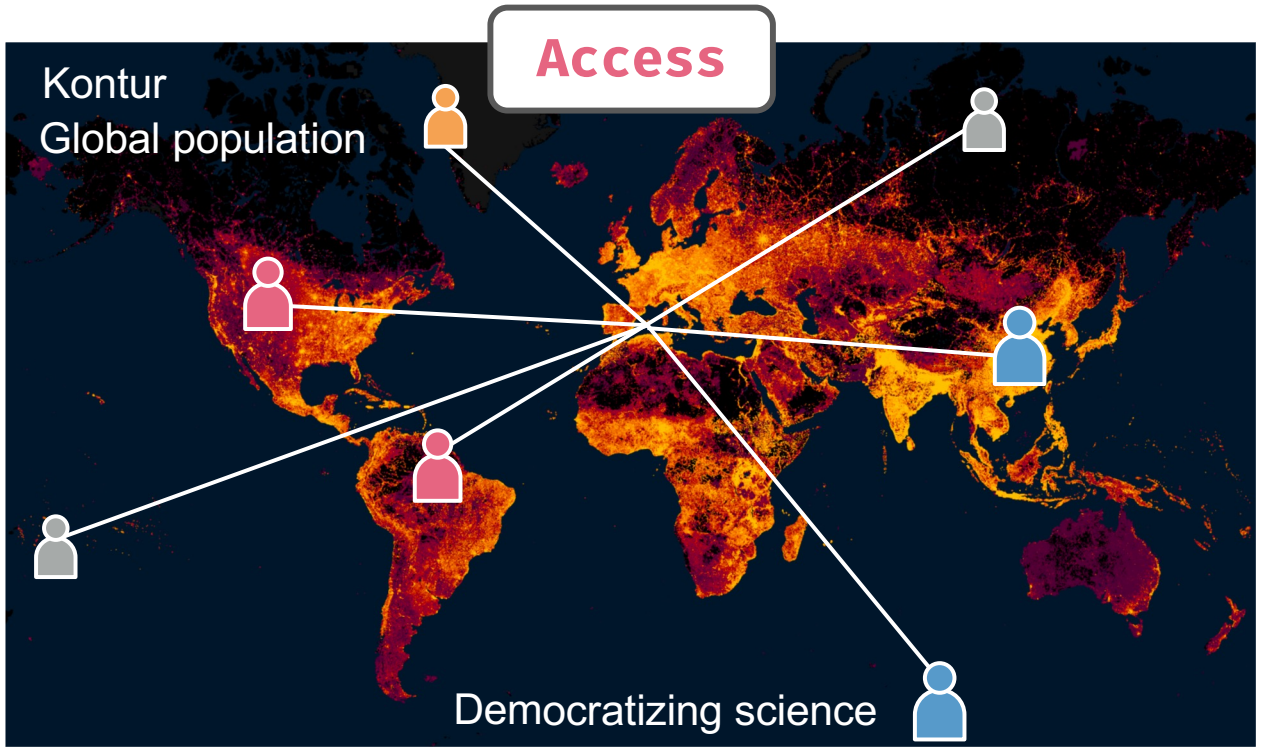
Easy to use, customizable – same software on your local/HPC/cloud

Collaboration made easy – co-coding, shared tools

Eliminates technology bottlenecks – shared challenges

No software expertise needed – cloud-computing as a service

Early careers, underrepresented groups, and non-R1 academic institution researchers benefit the most



CryoCloud helps accelerate science and makes open science easy

Data read-ins 1-2 orders of magnitude faster

Easy to use, customizable – same software on your local/HPC/cloud

Collaboration made easy – co-coding, shared tools

Eliminates technology bottlenecks – shared challenges

No software expertise needed – cloud-computing as a service

A consistent platform provides a place where individuals and communities can rapidly mobilize when need arises

- ICESat-2 Science Team pre-launch training: Dec 2022
- FOGGS: March 2023
- ICESat-2 Hackweek: Aug 2023
- WAIS: Sept 2023
- ICESat-2 Science Team Meeting: Oct 2023
- AGU Year of Open Science: Dec 2023

Norwegian ML Workshop (Apr), QGreenland Workshop (May), NSIDC User Working Group (Sept), GeoSmart Hackweek (Oct)

Cost to run a workshop in the cloud

Case study: QGreenland Workshop

Standing up own hub for 1 month

$4500 for hub

$75 in cloud credits

Science experts to advise in constraining hub needs

2-4 work weeks - Build and maintain hub

With CryoCloud

$75 in cloud credits

2 hours to 2 work days - Science and/or tech experts to advise on user & infrastructure needs

Shared resource community model – expertise and technology

QGreenland + CryoCloud

Cost to run a workshop in the cloud

Case study: QGreenland Workshop

Standing up own hub for 1 month	With CryoCloud
~~$4500 for hub~~	$75 in cloud credits
$75 in cloud credits	2 hours to 2 work days - Science and/or tech experts to advise on user & infrastructure needs
Science experts to advise in constraining hub needs	**Shared resource community model** – expertise and technology
~~2-4 work weeks - Build and maintain hub~~	

+

Cost to run a workshop in the cloud

Case study: QGreenland Workshop

Standing

$4500 for hu

$75 in cloud

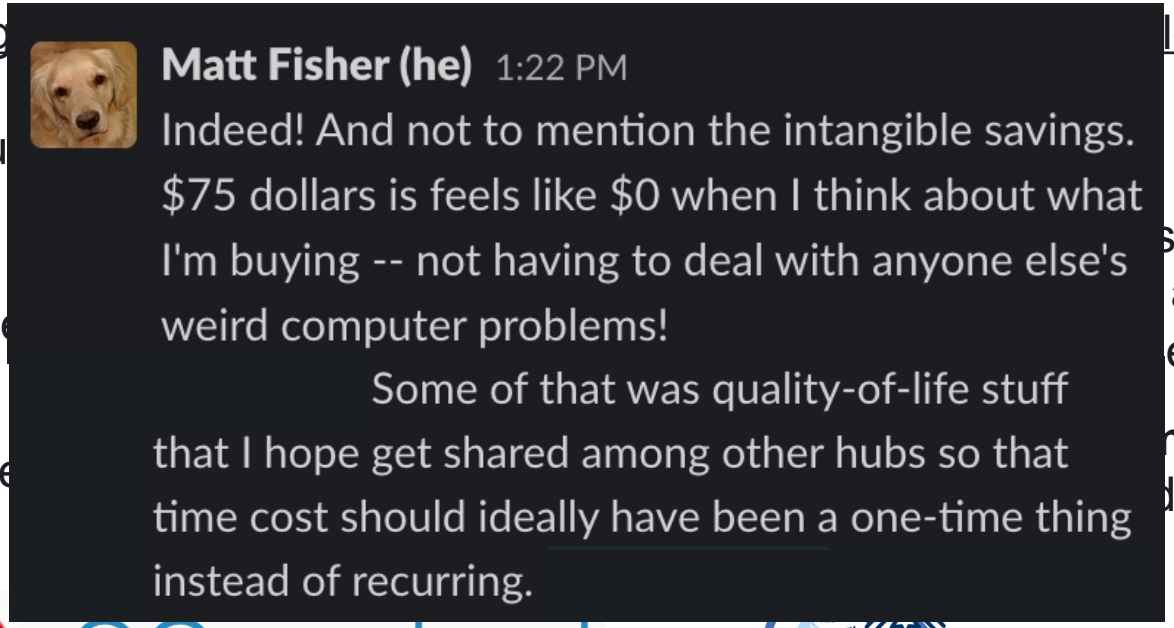
Science exp
hub needs

2-4 work wee

loud

s - Science
advise on
eeds

mmunity
d technology



QGreenland + CryoCloud

CryoCloud helps accelerate science and makes open science easy

Data read-ins 1-2 orders of magnitude faster

Easy to use, customizable – same software on your local/HPC/cloud

Collaboration made easy – co-coding, shared tools

Eliminates technology bottlenecks – shared challenges

No software expertise needed – cloud-computing as a service

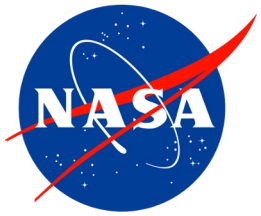
Driving social and technological innovation forward

Active engagement of scientific community in building new community models

Technological Innovation	Social Innovation
Technological knowledge	Management knowledge
R&D and ICT investments	Education and experience
Research and Development	Organisation, management, labour
Knowledge creation	Acquisition, integration, application of new knowledge
Explains 25% of innovation success	Explains 75% of innovation success

Source: Erasmus University: Competition and Innovation monitor (2006)

Open science as a process, not a product



cryointhecloud.com

Thank You!

tsnow03.github.io

@tsnow03 @TashaMSnow tsnow@mines.edu