EGI 2023, June 20 2023, Poznań, Poland
We did it! The innovation impact of EGI-ACE
Tina Odaka (IFREMER)

PANGEO

A European big data analytics platform in geosciences with Pangeo

Pangeo-EOSC with EGI

- What is Pangeo?
- Challenge
- Solution
- Service from EGI
- Impact
- Highlights
- Lessons learned and next step

A Culture of Collaboration

PANGEO

Lamont-Doherty Earth Observatory
Columbia University | Earth Institute

simula

cnes

Met Office

National Centre for Atmospheric Science
Natural Environment Research Council

climate extremes
ARC centre of excellence

ILLINOIS
University of Illinois at Urbana-Champaign

NOAA
National Oceanic and Atmospheric Administration
U.S. Department of Commerce

EOSC
NORDIC

UiO
University of Oslo

universität innsbruck

Ifremer

NCAR
National Center for Atmospheric Research

unidata

PRINCETON UNIVERSITY

British Antarctic Survey

USGS
science for a changing world

Element 84

UNIVERSITY OF OXFORD

LOPS

LEGOS

UNIVERSITY AT ALBANY
State University of New York

UNIVERSITY of WASHINGTON

Net Carbon

Centre for Environmental Data Analysis
Science and Technology Facilities Council
Natural Environment Research Council

IMAU
Institute for Marine and Atmospheric research Utrecht

INFORMATICS LAB

NVIDIA

ANACONDA.

Galaxy PROJECT

Inria

HARVARD UNIVERSITY

MIAMI

CSIRO

EOSC-PILLAR
Earth Science VRE

GEOMAR
Helmholtz Centre for Ocean Research Kiel

IGE

The Alan Turing Institute

Stockholms universitet

Massachusetts Institute of Technology

Los Alamos
NATIONAL LABORATORY
EST. 1943

UMBC
University of Maryland Baltimore County
1966

UNIVERSITY OF LEEDS

UNIVERSITY OF TORONTO

UNIVERSITY OF SASKATCHEWAN

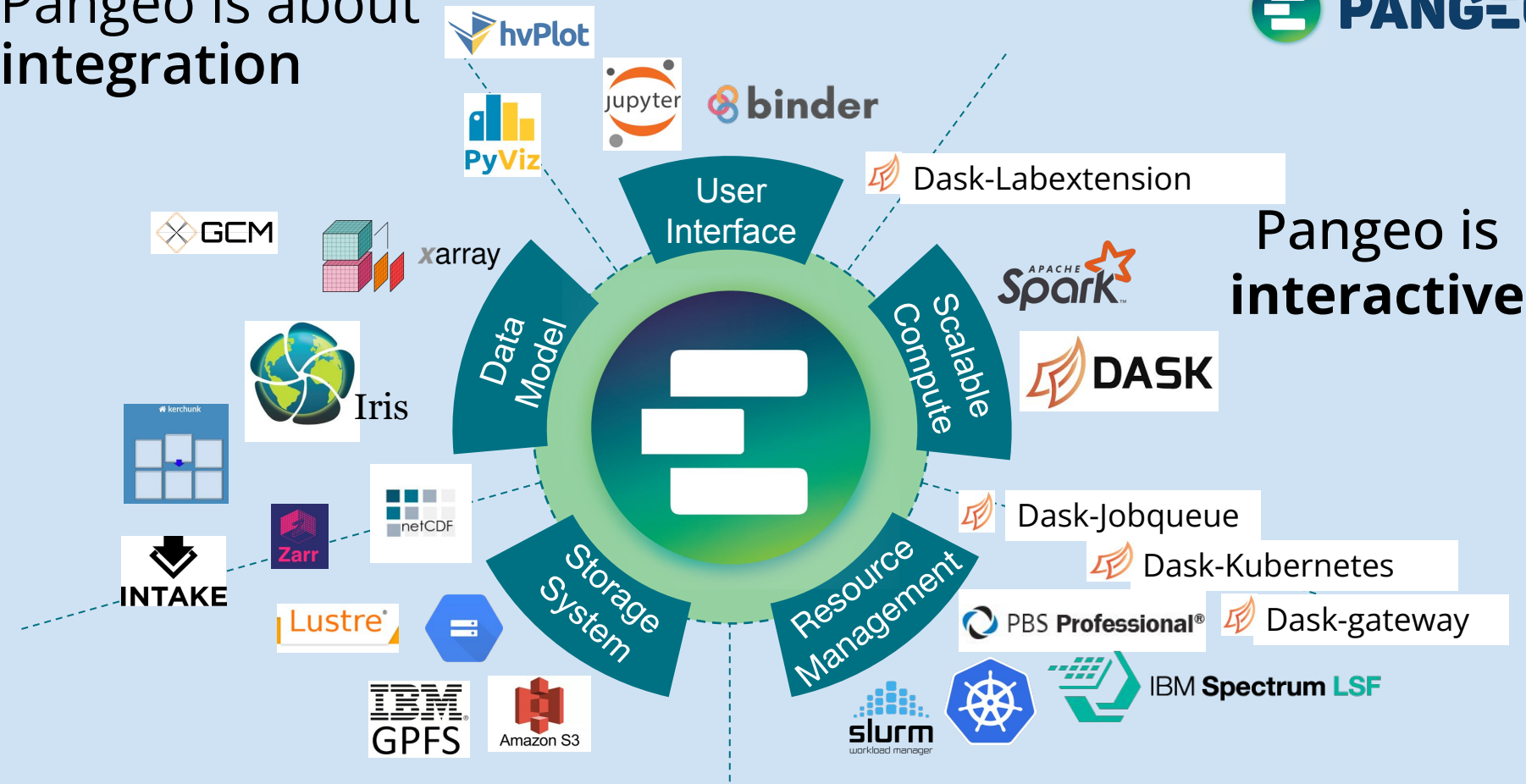
Moss Landing Marine Laboratories

MEMORIAL UNIVERSITY

Microsoft

Based on GitHub and papers affiliations

Pangeo is about **integration**



PANGEO

Pangeo is **interactive**

Pangeo platform *is* ***scalable*** *from laptop to cloud to HPC*

Pangeo European community

- Cloud
  - Commercial clouds (AWS, Microsoft, Google)
- Can’t we have European deployment ??

Pangeo European community

- Cloud
  - Commercial clouds (AWS, Microsoft, Google)
- Can’t we have European deployment ??

YES WE CAN!

thanks to European collaborations!

Pangeo & eosc

PANGEO

Custom environments

Online content

JB

RoHub

EOSC Community Hub

jupyterhub

xarray

pandas

NumPy

lab

jupyter

julia

Data science interface

Cloud infrastructure

C-SCALE

EGI-ACE

Data

egi

PANGEO

aws

CREODIAS

cesnet

Compute

ORCID

AUTHENTICATE

EOSC Authentication for students, researchers, data scientists, etc.

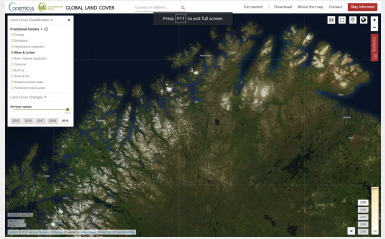
Contribution to Pangeo community from European side (US side: i2c2 Pangeo deployment)

Achievements thanks to EGI

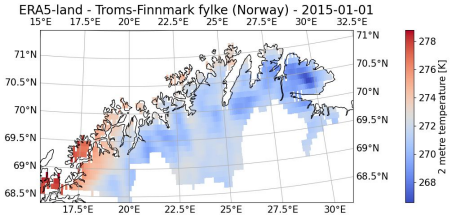
- Pangeo Reliance: Arctic browning and reindeer husbandry in a changing climate
- Pangeo-Julia: Recurrence Quantification Analysis for forest change detection
- Pangeo Training Infrastructure As a Service (PTIAsS)

Pangeo-RELIANCE
Use case with data analytics capabilities

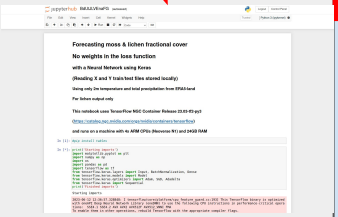
Arctic browning and reindeer husbandry in a changing climate: forecasting moss & lichen fractional cover with machine learning required computational capabilities, including GPUs that the RELIANCE project was not able to cover with EGI notebooks.



Copernicus Global Land Cover provides 100m resolution maps of cover fraction expressed in % ground cover per pixel for 10 base classes including moss & lichen.



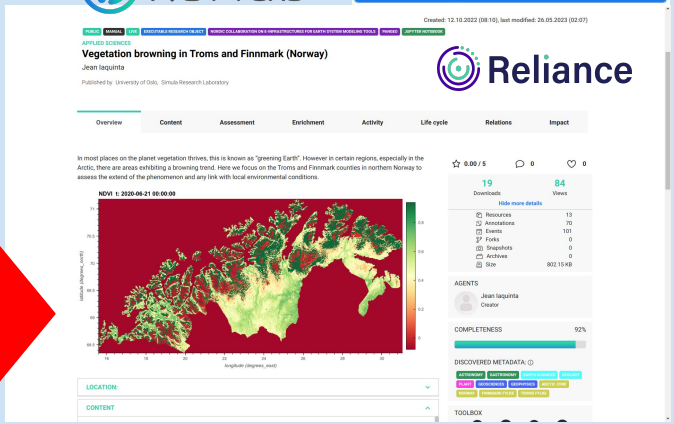
ERA5-land is a meteorological reanalysis dataset providing a consistent view of the evolution of land variables (here 2m air temperature) at an horizontal resolution of 0.1°× 0.1° (native 9 km).



EGI-ACE

RoHub

EXECUTABLE RESEARCH OBJECT



Binder

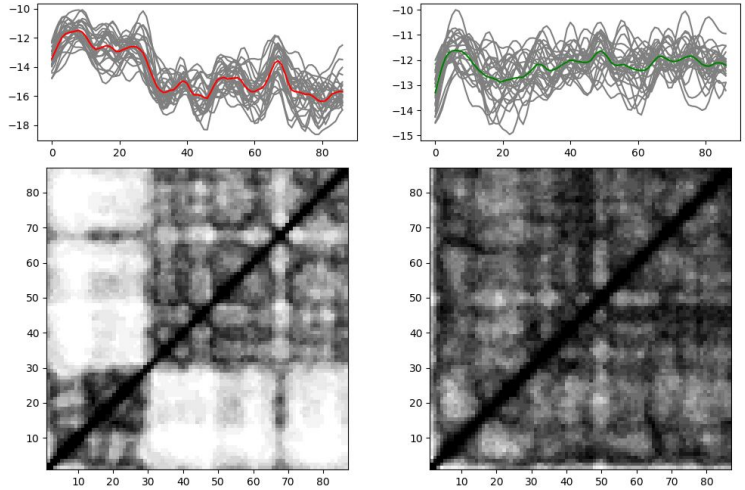
binder + jupyterhub

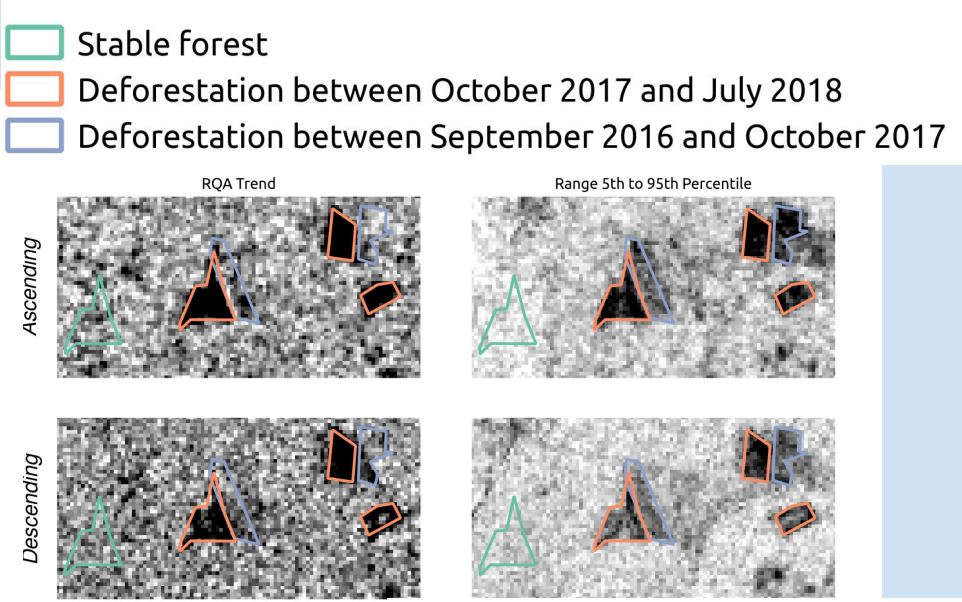
Use pre-processed and trained neural network that do not require too much resources.

Pangeo-Julia

- Use Recurrence Quantification Analysis for forest change detection
- Sentinel-1 data
- Scale this analysis to larger areas Starting with Central Germany and Yucatan;

- Testing of the JuliaDataCubes ecosystem on cloud computing infrastructure;
- Scalability tests of the JuliaDataCubes packages;
- Integration of the JuliaDataCubes Ecosystem into Pangeo





Pangeo@EOSC for TIAaS

- **One framework** for Pangeo@EOSC: fully reproducible and easy to maintain;
- **On-demand** deployments for teaching (and research) purposes;
- Compute and storage resources customised depending on needs:
  - FOSS4G workshop (August 2022);
  - CLIVAR/IASC CMIP6 Bootcamp for young researchers (from October 2022);
  - eScience Course for master and PhD students (November 2022);
  - Worldwide Climate Informatics Reproducible challenge (May 2023);
- **More to come!**
  - BiDS23: one Pangeo-Julia tutorial and Pangeo-Python tutorial, Pangeo-OpenEO tutorial (*submitted*)!

Pangeo: where can I find information?

- Presentation by Jean tomorrow, wednesday 12:40 - (Dublin)
  - "Data analytics platforms, tools and VREs for EOSC" session
  - “Open big data geoscience with Pangeo@EOSC”
- http://pangeo.io/index.html#what-is-pangeo
  - (https://github.com/pangeo-data/pangeo)
- Try some examples in the Pangeo Gallery
  - http://gallery.pangeo.io
- Subscribe in discourse, start discussions
  - https://discourse.pangeo.io/
- Follow pangeo-data from Gitlab & Gitter
- Interesting informations exist also in blogs
  - https://medium.com/pangeo
- Code of conduct.
  - https://github.com/pangeo-data/governance/blob/master/conduct/code_of_conduct.md