

A cyber cowboy, a humanoid robot with a cowboy hat and a dark, metallic suit, stands in a western-themed setting. The robot is positioned in the foreground, looking out a large window. Outside the window, a steam locomotive is visible in a hazy, sunlit landscape. The robot's hands are resting on a wooden table in the foreground, which holds a glass bottle and some mechanical components. The overall atmosphere is a blend of the Old West and futuristic technology.

Cyber cowboys: wrangling big data on the open science frontier

Tyson Lee Swetnam
Texas A&M Institute for Data Science
2023-04-03



Overview

The Big Data Landscape

2023 Year of Open Science

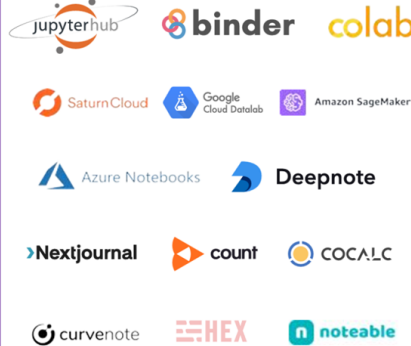
Becoming Cloud-Native

AI in the workplace

CyVerse Science Use Cases

MACHINE LEARNING & ARTIFICIAL INTELLIGENCE

DATA SCIENCE NOTEBOOKS



DATA SCIENCE PLATFORMS



ML PLATFORMS



OPEN SOURCE

STAT TOOLS & LANGUAGES



ML OPS & INFRA



The Big Data Landscape is really, really, really, big

- >100 Thousand data scientists in the US¹
- ~26.9 Million software engineers in the world²
- >15 Billion devices connected to the Internet³
- ~7.9 Zettabytes (7.9 Billion Terabytes) of data on the internet⁴

1 <https://www.bls.gov/ooh/math/data-scientists.htm>

2 https://en.wikipedia.org/wiki/Software_engineering_demographics

3 <https://www.statista.com/statistics/1290925/internet-access-by-device-us/>

4 <https://firstsiteguide.com/big-data-stats/>



AUGUST 25, 2022

OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay



› OSTP

› BRIEFING ROOM

› PRESS RELEASES

Source: <https://www.whitehouse.gov/ostp/news-updates/2022/08/25/ostp-issues-guidance-to-make-federally-funded-research-freely-available-without-delay/>



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

August 25, 2022

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM:

Dr. Alondra Nelson


A handwritten signature in blue ink, reading "Alondra Nelson".

Deputy Assistant to the President and Deputy Director for Science and Society
Performing the Duties of Director
Office of Science and Technology Policy (OSTP)

SUBJECT: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

This memorandum provides policy guidance to federal agencies with research and development expenditures on updating their public access policies. In accordance with this memorandum, OSTP recommends that federal agencies, to the extent consistent with applicable law:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.



How many of you are trained to make data accessible immediately after you analyze it?

document

curate
share

Capacity vs Capability



Compute
Storage
Data
Software



Digital Literacy
Data Science Skills



JANUARY 11, 2023

FACT SHEET: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research



› [OSTP](#)

› [BRIEFING ROOM](#)

› [PRESS RELEASES](#)

*OSTP launches Year of Open Science to advance national open science policies
across the federal government in 2023*

Source: <https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/>

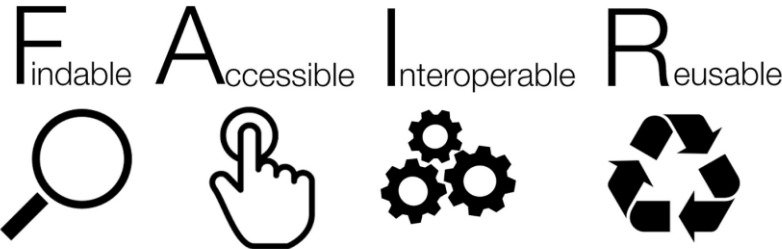
2023 - Federal Year of Open Science Initiative



National Institutes
of Health



What is Open Science?





Home

Before FOSS Starts

Schedule

Code of Conduct

Glossary & Acronyms



Welcome to Foundational Open Science Skills
(FOSS) Spring 2023!

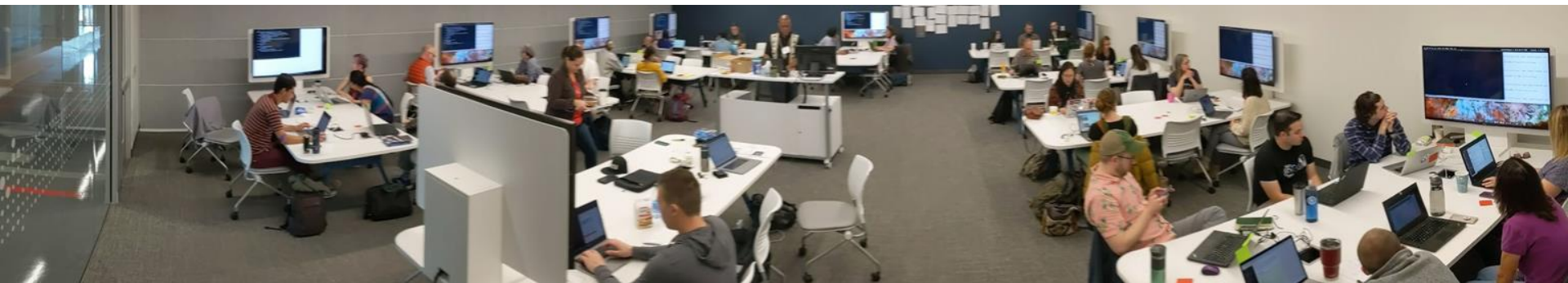


On this page

Workshop Structure

Capstone Project

Expected Outcomes



<https://foss.cyverse.org>

**Lessons**

0. The Shell and Git

1. Open Science

2. Project Management

3. Managing Data

4. Documentation and
Communication

5. Version Control

6. Reproducibility I: Repeatability

7. Reproducibility II: Containers

Introduction to Open Science

**✓ Learning Objectives**

After this lesson, you should be able to:

- Explain what Open Science is
- Explain the components of Open Science
- Describe the behaviors of Open Science
- Explain why Open Science matters in education, research, and society
- Understand the advantages and the challenges to Open Science
- Identify who the practitioners of Open Science are
- Understand the underlying Ethos of Open Science

On this page

What is Open Science?

Open Access Publications

Open Data

Open Educational Resources

Open Methodology

Open Peer Review

Open Source Software

Breakout Discussion

Components of Open Science

WHY do Open Science?

Ethos of Open Science

Recommended Open Science
Communities

Self Assessment

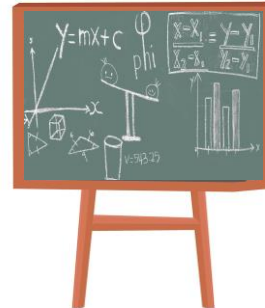
What is Open Science?

If you ask a dozen researchers this question, you will probably get just as many answers.

This means that Open Science isn't necessarily a set of checkboxes you need to tick, but rather a holistic approach to doing science.

<https://foss.cyverse.org>

Personal Computing



Personal Computing

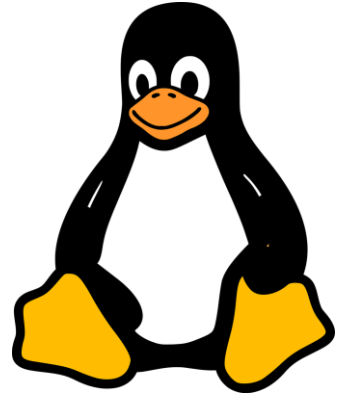


76%



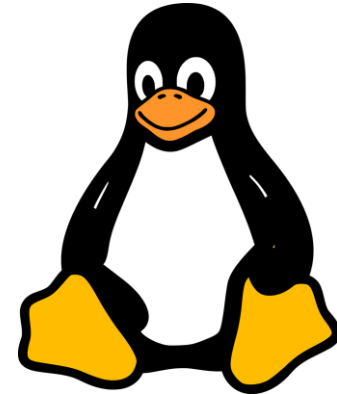
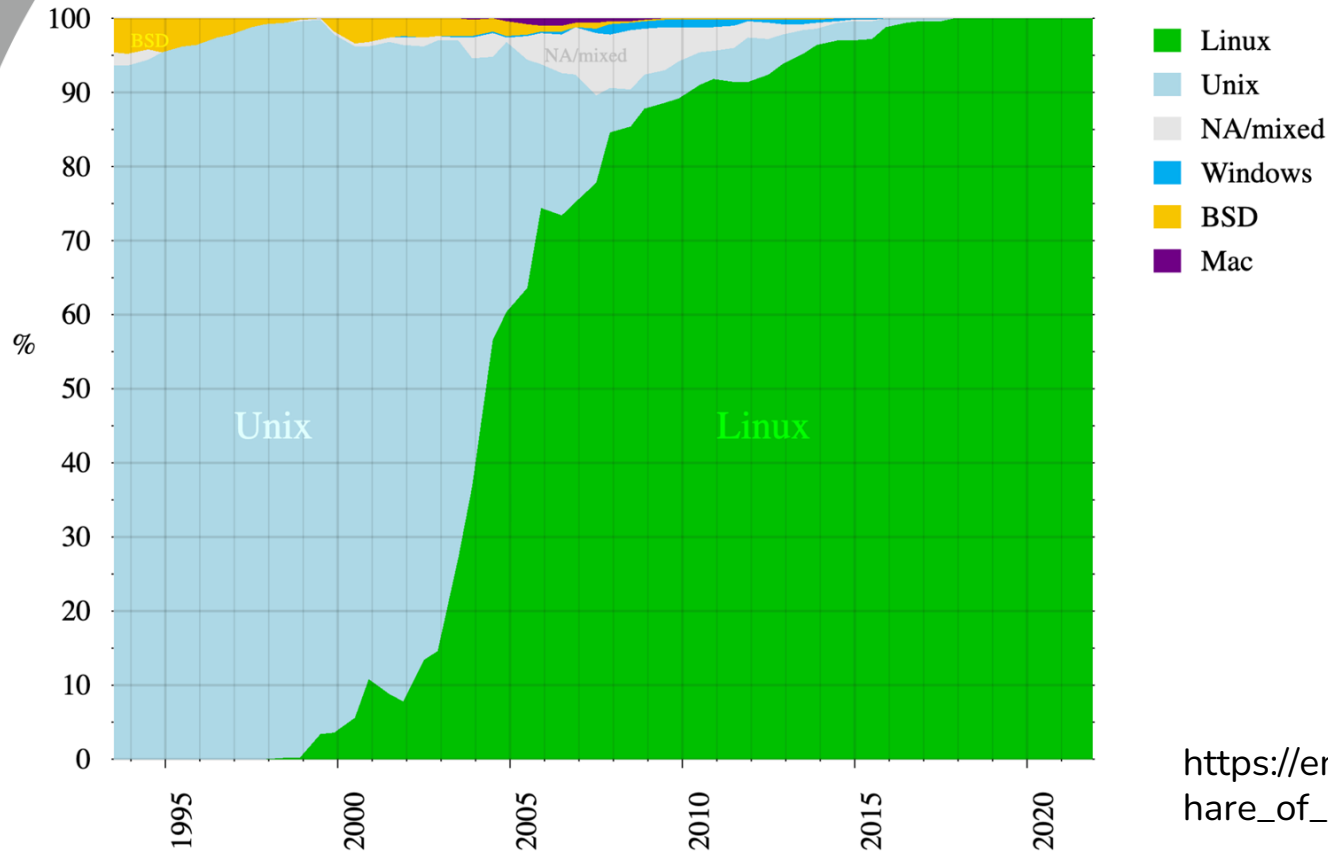
macOS

16%



5%

Cloud Computing & SuperComputing

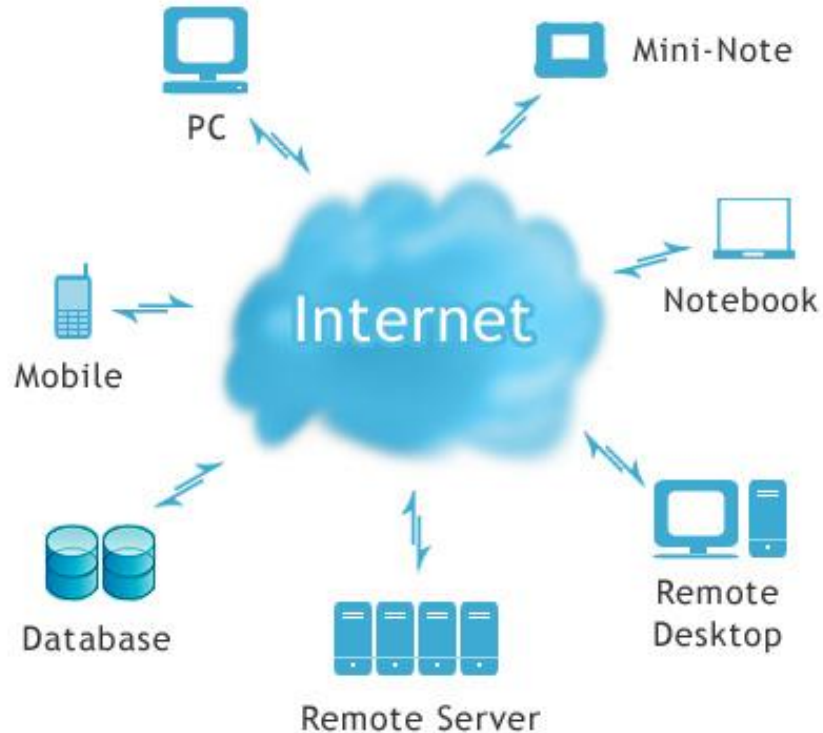


https://en.wikipedia.org/wiki/Usage_share_of_operating_systems

Let's all use "The Cloud!"



Let's all use "The Cloud!"



Let's all use “The Cloud!”

“Free” Cloud services are available almost everywhere for academic educators & researchers

“Starter Tiers” can be good for teaching and getting a feel for things, but in general are not useful for big data analyses



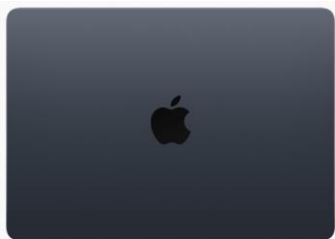
My Daily Drivers



Laptop

Desktop

HPC & Data Center



arm

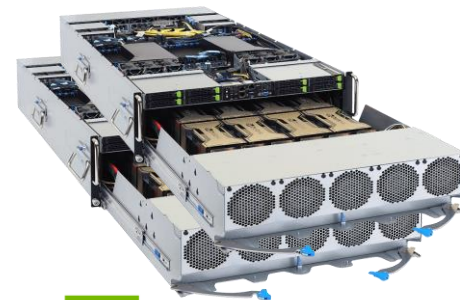


macOS



AMD

ubuntu



AMD



Rocky Linux™

My Daily Drivers



GitHub



OpenAI



Microsoft

CONDA

YAML

python
powered
`print("Hello, world!")`

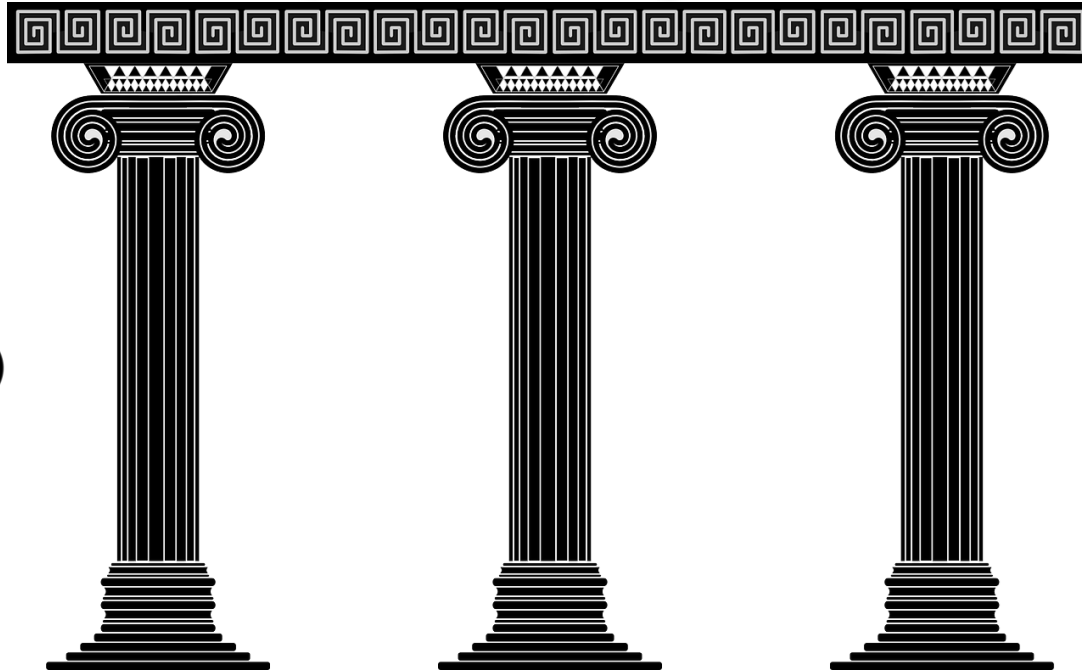
JavaScript



Data
Proximate
Computing

Dynamic
Elastic
Scaling

Analysis
Ready
Data



<https://doi.org/10.1109/MCSE.2021.3059437>

Ryan Abernathey's 2022 PANGEO talk at OpenOceanCloud: <https://vimeo.com/670782104>

Data Proximate Computing



Mainframe

Terminal



Desktop

High Density Servers

Cloud-enabled Apps



1980s



Many Servers

Web Hosting



2020+

1970s

1990s-2010s

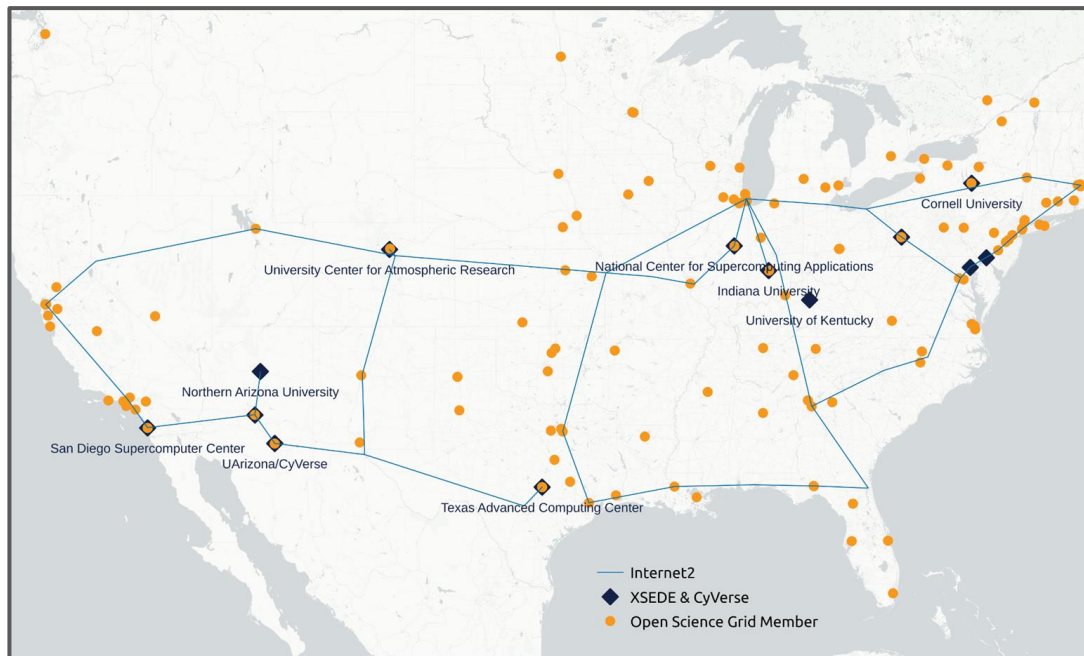
Data Proximate Computing

Connect to Public Cyberinfrastructure & Commercial Cloud via Internet2

ADVANCE TO
ACCESS

MAKING A SUCCESSFUL TRANSITION

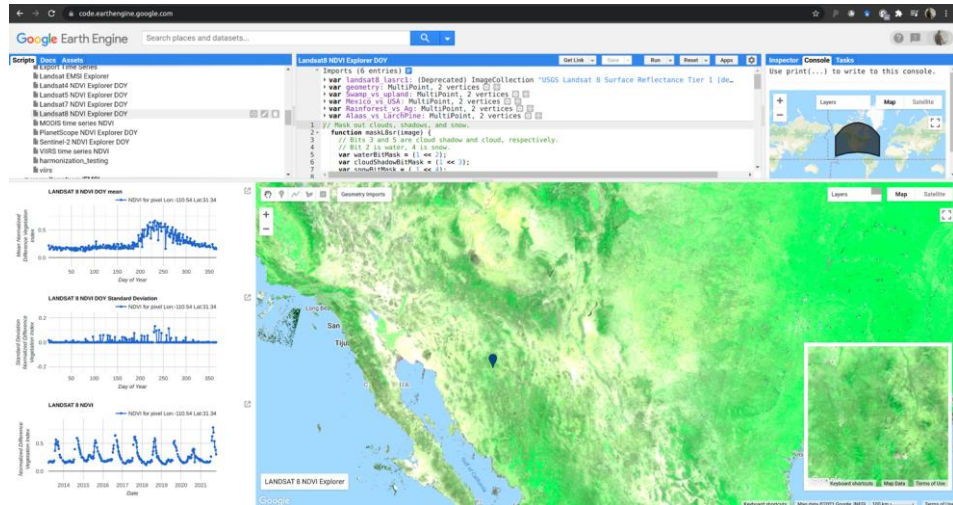
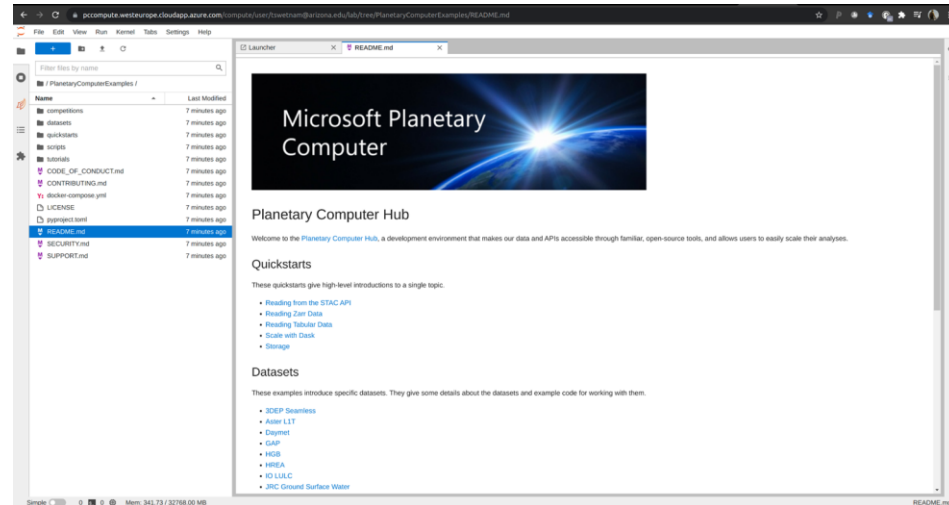
PATH PARTNERSHIP to ADVANCE
**THROUGHPUT
COMPUTING**



Dynamic and Elastic Scaling

<https://planetarycomputer.microsoft.com/>

<https://code.earthengine.google.com/>

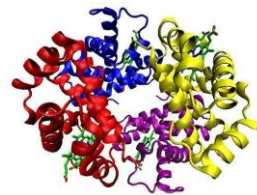
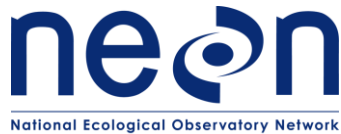


<https://github.com/microsoft/PlanetaryComputerExamples>

<https://github.com/giswqs/Awesome-GEE>



Analysis Ready Data



Analysis Ready Data

All major scientific data archives are hosted on the cloud - most are free

Programmatically scalable to millions of concurrent requests

Metadata & indexing allow for ins

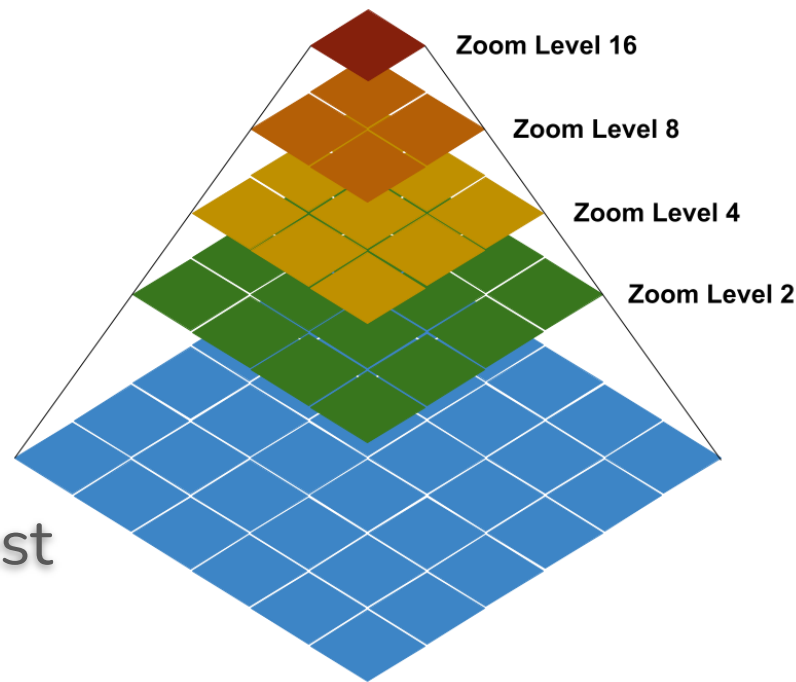
The screenshot shows the EarthData Search web application. On the left is a sidebar with filters for Collections, Categories, Features, Keywords, Platforms, Instruments, Organizations, Projects, Processing Levels, Data Format, Tiling System, Horizontal Data Resolution, and Additional Filters. The main area displays a map of the world with a search result for 'U.S. Landsat 4-8 Analysis Ready Data (ARD) V1'. The result shows 3,687,802 granules and a date range from 1982-11-01 to ongoing. A description states that ARD is consistently processed to the highest scene file standards and is ready for direct use in modeling and assessing landscape change. A 'View' button is visible next to the result. The bottom of the page shows a search bar and navigation links.

The screenshot shows the Planetary Computer Data Catalog web application. The header includes the Microsoft logo and navigation links for Planetary Computer, Explore, Data Catalog, Hub, Applications, Documentation, and a 'Request access' button. The main content area is titled 'Data Catalog' and includes a 'Filter by tags' input field. Below this, it states 'Datasets available through the Planetary Computer API' and provides a brief description of the API. Two satellite images are displayed: 'Landsat 8 Collection 2 Level-2' on the left and 'Sentinel-2 Level-2A' on the right. The bottom of the page shows the search bar and navigation links.

Analysis Ready Data

Why?

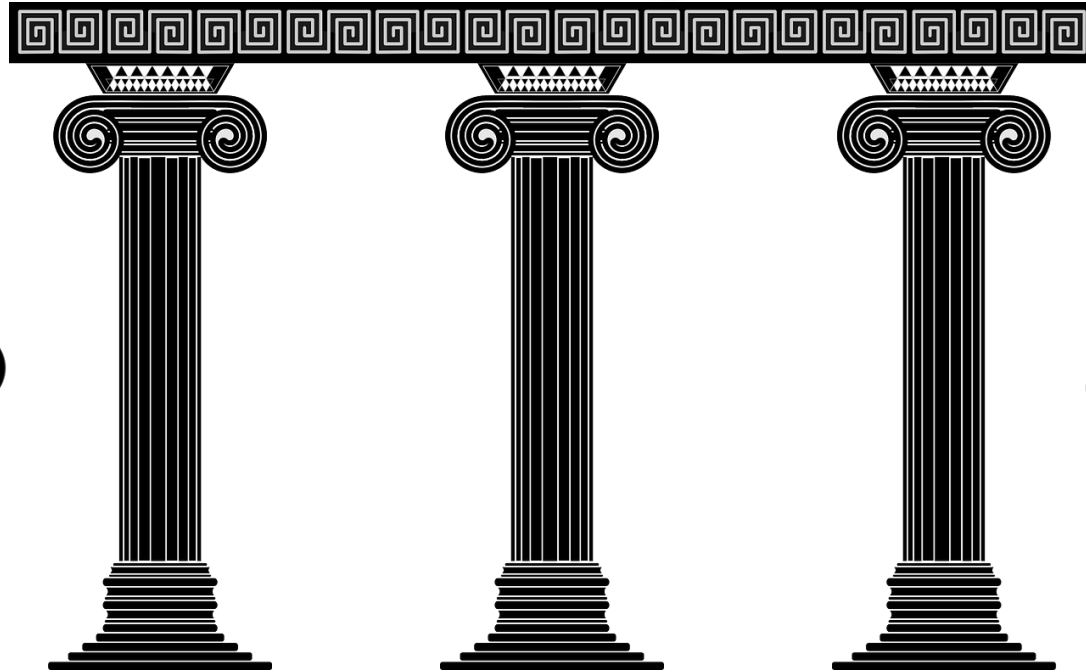
- Accessibility
- Preview without downloading
- Cloud Optimized
 - Byte range requests
 - Download only area of interest



Data
Proximate
Computing

Dynamic
Elastic
Scaling

Analysis
Ready
Data



<https://doi.org/10.1109/MCSE.2021.3059437>

Ryan Abernathey's 2022 PANGEO talk at OpenOceanCloud: <https://vimeo.com/670782104>

Artificial Intelligence & Machine Learning



AI investments are expected to go from
10 billion \$ to >200 billion \$ by 2025



SIGN IN NPR SHOP DONATE

NEWS CULTURE MUSIC PODCASTS & SHOWS SEARCH

TECHNOLOGY

Google shares drop \$100 billion after its new AI chatbot makes a mistake

February 9, 2023 · 10:15 AM ET



By Emily Olson



Shares for Google's parent company, Alphabet, dropped 9% Wednesday after its AI chatbot, Bard, gave an incorrect answer.

Artificial Intelligence & Machine Learning

+ New chat

Maize Phenotypic Vari  

Terraform Multi-File Configur


Convert TSV to Neo4J


Jupyter Notebook MidJourne


Load TSV file into Neo4J.


SSH key creation script.


Streamlit App with Autoscalin

 Clear conversations


 Dark mode

 My account


 Updates & FAQ



 Log out

Model: GPT-4



Design a research study which examines the phenotype of ten maize genetic lines. Describe a field trial where each of the ten lines is planted in replicated plots, estimate how many plants in each plot are required to assess the range of phenotypic variation under normal growing conditions in both an outdoor field in Iowa, and an indoor climate controlled greenhouse.





Title: Phenotypic Variation Assessment of Ten Maize Genetic Lines under Indoor and Outdoor Growing Conditions  

Objective: To examine the phenotypic variation in ten different maize genetic lines grown under normal field conditions in Iowa and in a climate-controlled greenhouse.

Study Design:

1. Selection of ten maize genetic lines: Choose ten diverse maize genetic lines that are representative of the species, including lines with known variations in phenotypic traits such as plant height, leaf size, kernel color, and disease resistance.

2. Experimental setup: The field trial will be conducted as a randomized complete block design (RCBD) with three repli  Regenerate response genetic lines in both outdoor



Artificial Intelligence & Machine Learning

New chat

"Field, University, City"

Maize Phenotypic Variation S

Terraform Multi-File C

Convert TSV to Neo4J

Jupyter Notebook MidJourne

Load TSV file into Neo4J.

SSH key creation script.

Clear conversations

Dark mode

My account

Updates & FAQ

Log out



Write a Python script to gather user input and generate the terraform.tfvars file



Here is a Python script that gathers user input and generates the `terraform.tfvars` file:



python

Copy code

```
import os
import subprocess

def main():
    # Gather user input
    openrc_file = input("Enter the path to your openrc.sh file: ")
    vm_count = int(input("Enter the number of virtual machines: "))
    image_name = input("Enter the image name: ")
    flavor_name = input("Enter the flavor name: ")
    key_pair = input("Enter the key pair name: ")
    storage_size = int(input("Enter the storage volume size (in GB): "))
    docker_image = input("Enter the Docker JupyterHub image name: ")

    # Load openrc.sh environment variables
    load_openrc(openrc_file)
```

Regenerate response



|



GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou¹, Sam Manning^{1,2}, Pamela Mishkin*¹, and Daniel Rock³

¹OpenAI

²OpenResearch

³University of Pennsylvania

March 20, 2023

Abstract

We investigate the potential implications of Generative Pre-trained Transformer (GPT) models and related technologies on the U.S. labor market. Using a new rubric, we assess occupations based on their correspondence with GPT capabilities, incorporating both human expertise and classifications from GPT-4. Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted. The influence spans all wage levels, with higher-income jobs potentially facing greater exposure. Notably, the impact is not limited to industries with higher recent productivity growth. We conclude that Generative Pre-trained Transformers exhibit characteristics of general-purpose technologies (GPTs), suggesting that as these models could have notable economic, social, and policy implications.

<https://arxiv.org/abs/2303.10130>

Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence

Shakked Noy Whitney Zhang
MIT MIT

March 2, 2023
Working Paper (not peer reviewed)

Abstract

We examine the productivity effects of a generative artificial intelligence technology—the assistive chatbot ChatGPT—in the context of mid-level professional writing tasks. In a preregistered online experiment, we assign occupation-specific, incentivized writing tasks to 444 college-educated professionals, and randomly expose half of them to ChatGPT. Our results show that ChatGPT substantially raises average productivity: time taken decreases by 0.8 SDs and output quality rises by 0.4 SDs. Inequality between workers decreases, as ChatGPT compresses the productivity distribution by benefiting low-ability workers more. ChatGPT mostly substitutes for worker effort rather than complementing worker skills, and restructures tasks towards idea-generation and editing and away from rough-drafting. Exposure to ChatGPT increases job satisfaction and self-efficacy and heightens both concern and excitement about automation technologies.

https://economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1.pdf

The Impact of AI on Developer Productivity: Evidence from GitHub Copilot

Sida Peng,^{1*} Eirini Kalliamvakou,² Peter Cihon,² Mert Demirer³

¹Microsoft Research, 14820 NE 36th St, Redmond, USA

²GitHub Inc., 88 Colin P Kelly Jr St, San Francisco, USA

³MIT Sloan School of Management, 100 Main Street Cambridge, USA

*To whom correspondence should be addressed; E-mail: sidpeng@microsoft.com.

Abstract

Generative AI tools hold promise to increase human productivity. This paper presents results from a controlled experiment with GitHub Copilot, an AI pair programmer. Recruited software developers were asked to implement an HTTP server in JavaScript as quickly as possible. The treatment group, with access to the AI pair programmer, completed the task 55.8% faster than the control group. Observed heterogeneous effects show promise for AI pair programmers to help people transition into software development careers.

<https://arxiv.org/abs/2302.06590>

Artificial Intelligence & Machine Learning



GPT v4

(Generative Pre-trained Transformer)



Hugging Face

LLM

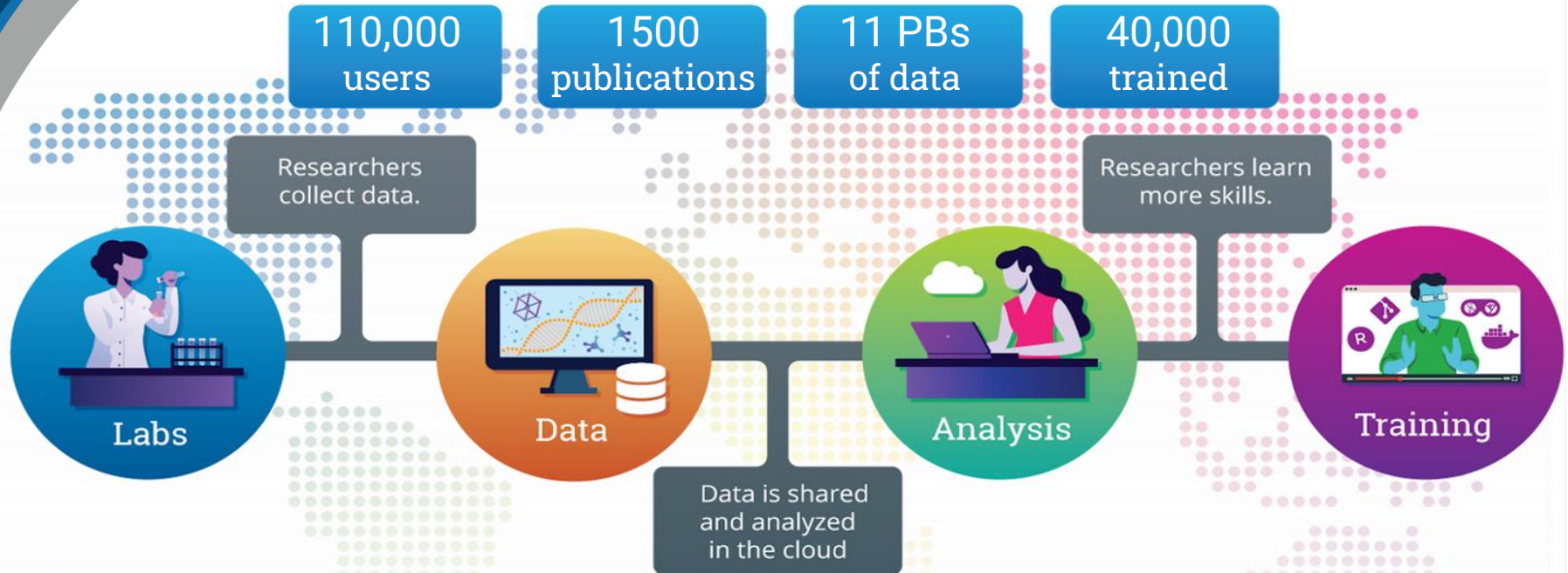
(Large Language Model)

- Use well defined, progressive, prompts
- Create or refine code you are struggling with
- Ask it about a specific piece of code or error you cannot understand
- For tasks that are otherwise a drudgery & / or subject to human error



The Open Science Workspace

DBI-0735191, DBI-1265383, DBI-1743442 and OAC-1664172



CYVERSE® Professional

- Federation with local and commercial cloud and high-performance computing
- Integration with local user identity management systems
- Security compliance

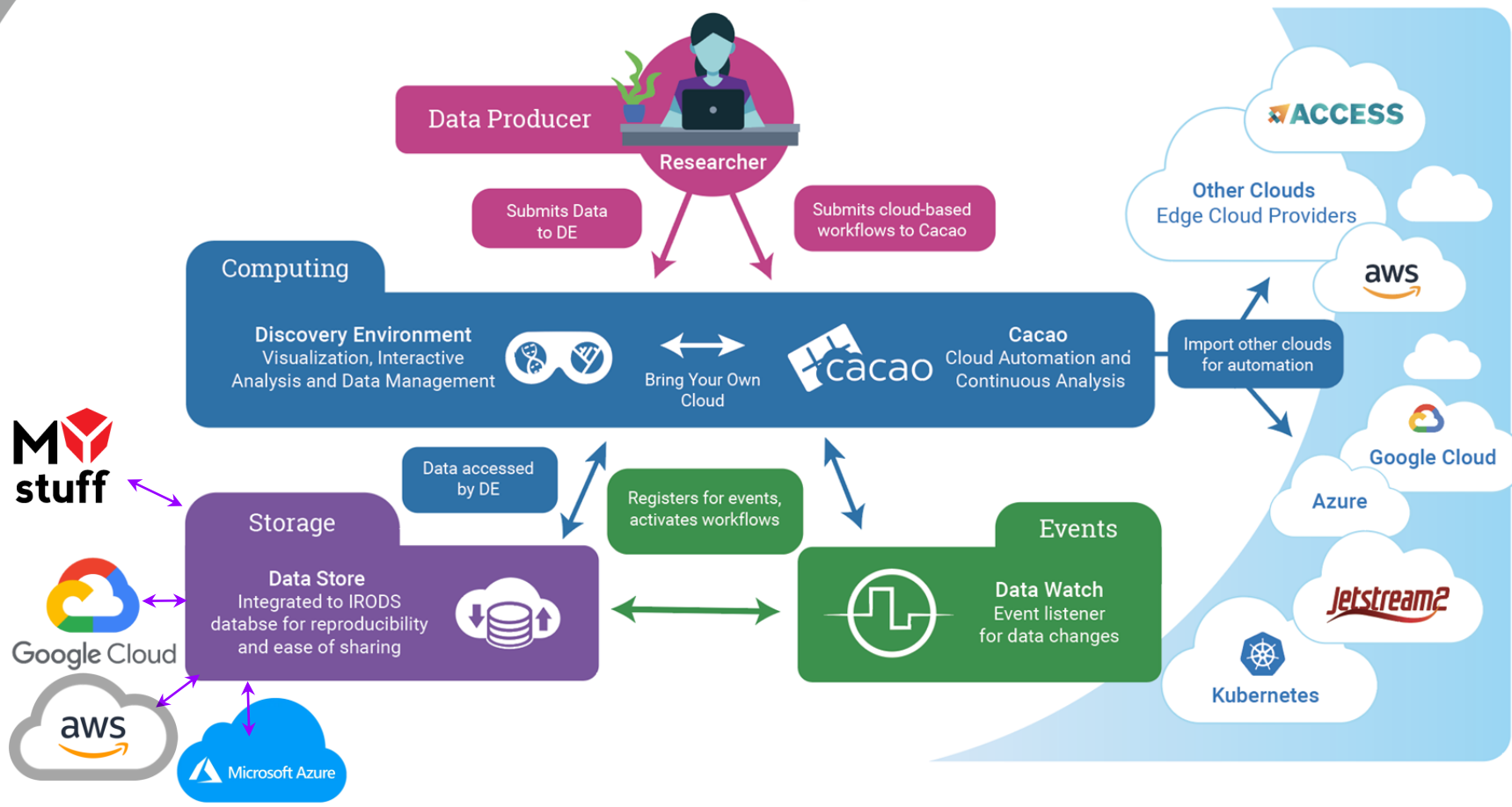
CYVERSE® Health

- Secure Perimeter (VPN)
- HIPAA Compliant Platforms inside of Perimeter
- XNAT
- Training

CYVERSE® Defense

- ITAR Compliance
- Receive data from multiple sensors
- Policy based data visibility and sharing
- Support for multiple teams and data partitioning

How does CyVerse work?



CyVerse Collaborations

 **ACCESS**

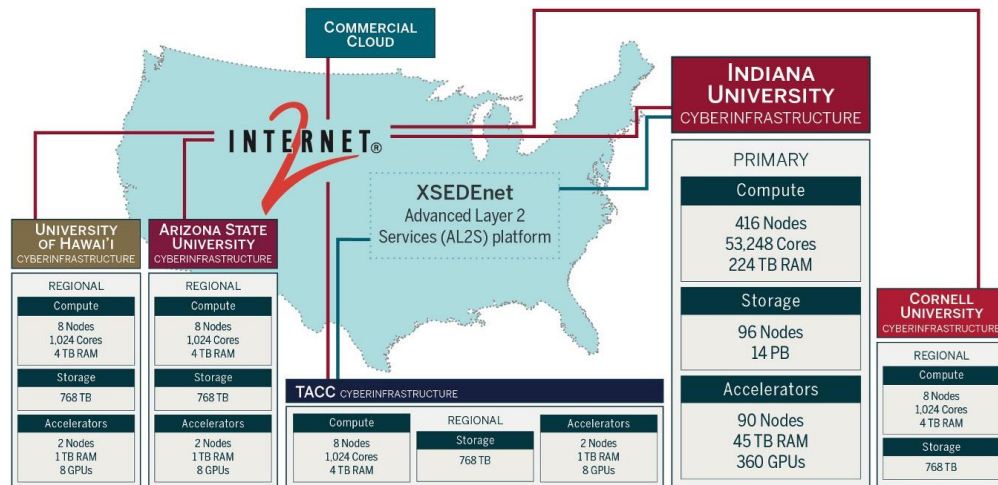
Jetstream2



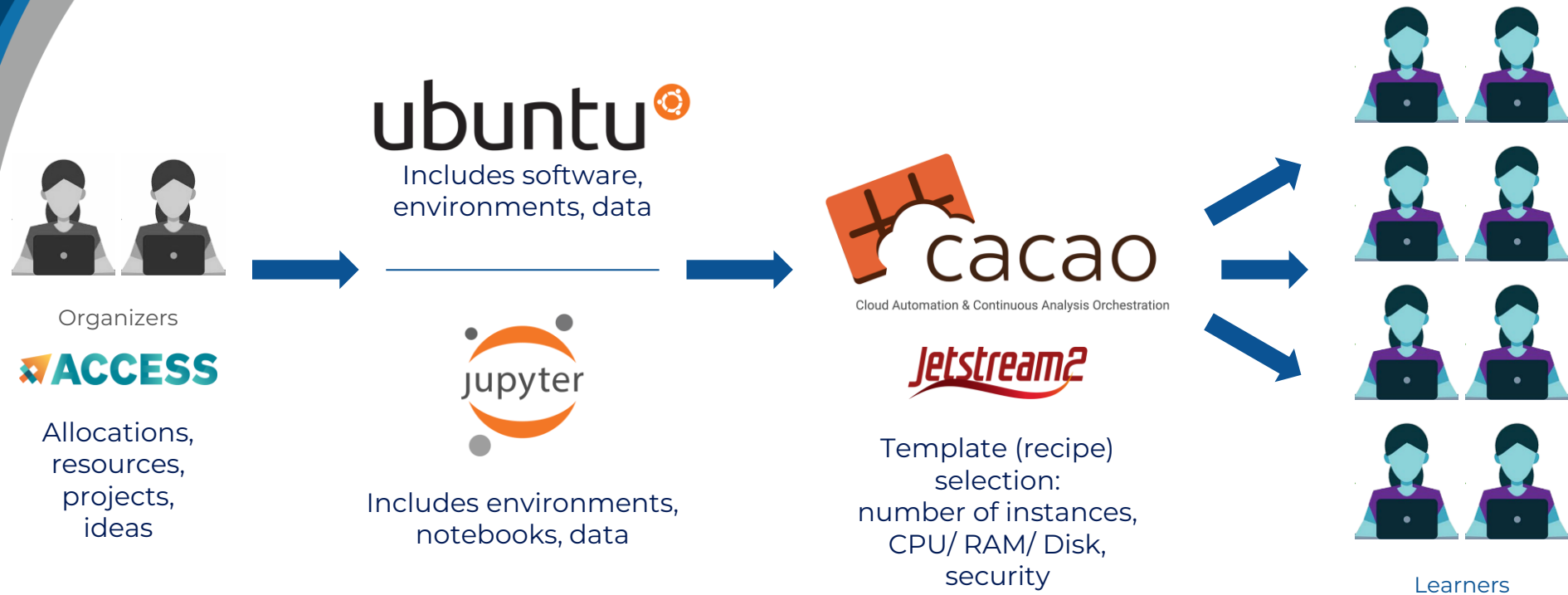
Cloud Automation & Continuous Analysis Orchestration



Award # [OAC 2005506](#)



CyVerse Collaborations



CyVerse Use Case # 1

Transforming a Quarter Petabyte of Field Phenomics Data Into Functional Traits

emmanuelgonzalez@arizona.edu

Emmanuel Gonzalez

*Duke Pauli Lab
Ph.D. Candidate
University of Arizona*



Pauli Lab



Collecting data is becoming easier and cheaper

Robots



Carts



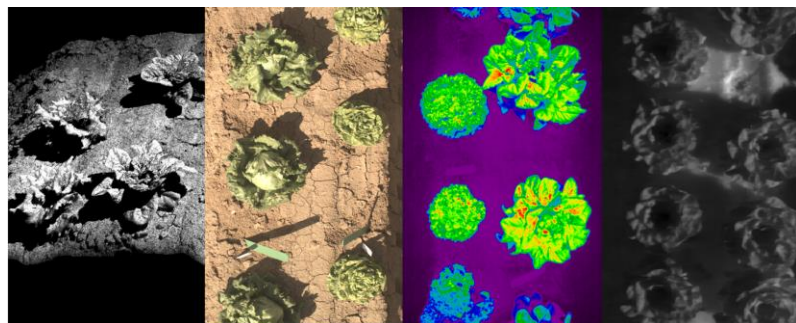
Drones



Phones



World's Largest Field Robot

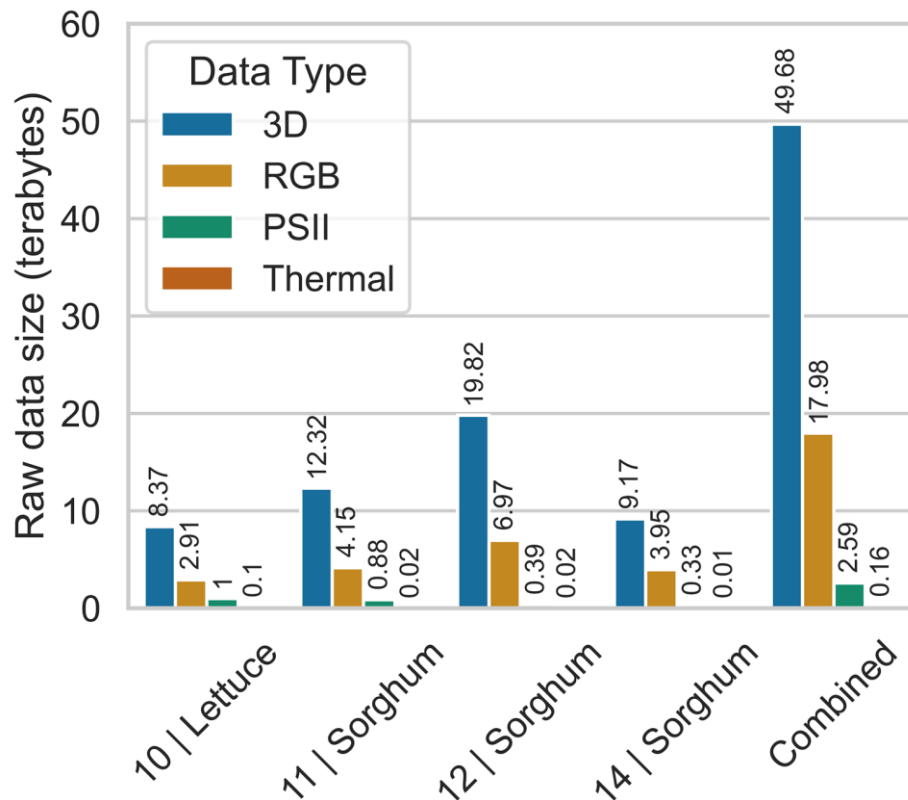


3D

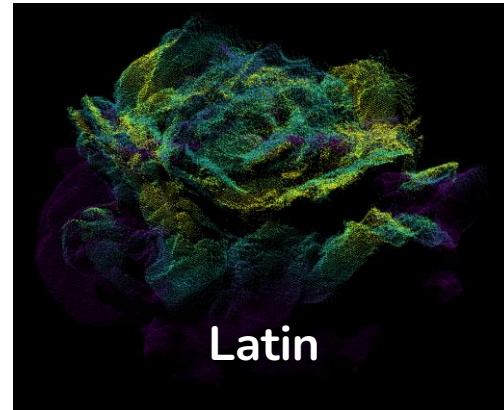
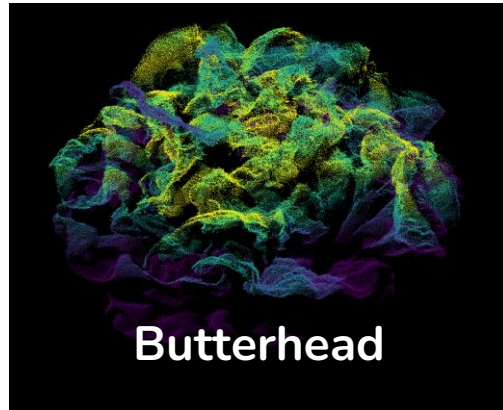
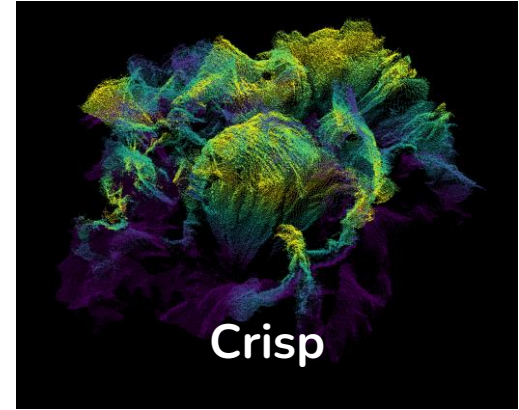
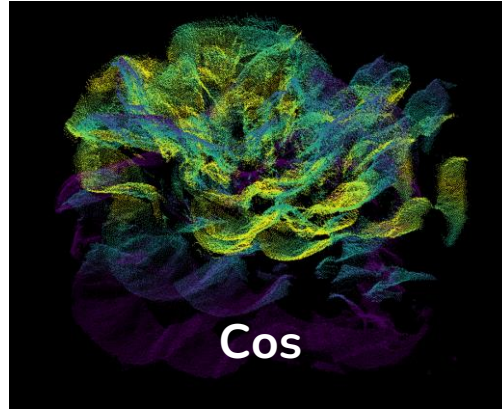
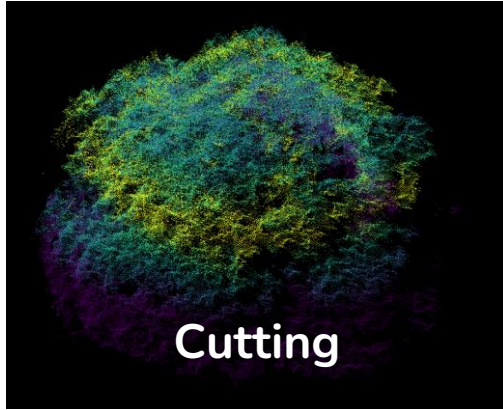
RGB

Fluorescence

Thermal



Lettuce exhibits large amounts of phenotypic variation



Goal: repeatable, fine-scale phenotype extraction

3D

20,000 plants * 32 scans * 4 phenotypes
= 2.6M individual plant phenotypes

RGB

20,000 plants * 36 scans * 1 phenotype
= 720K individual plant phenotypes

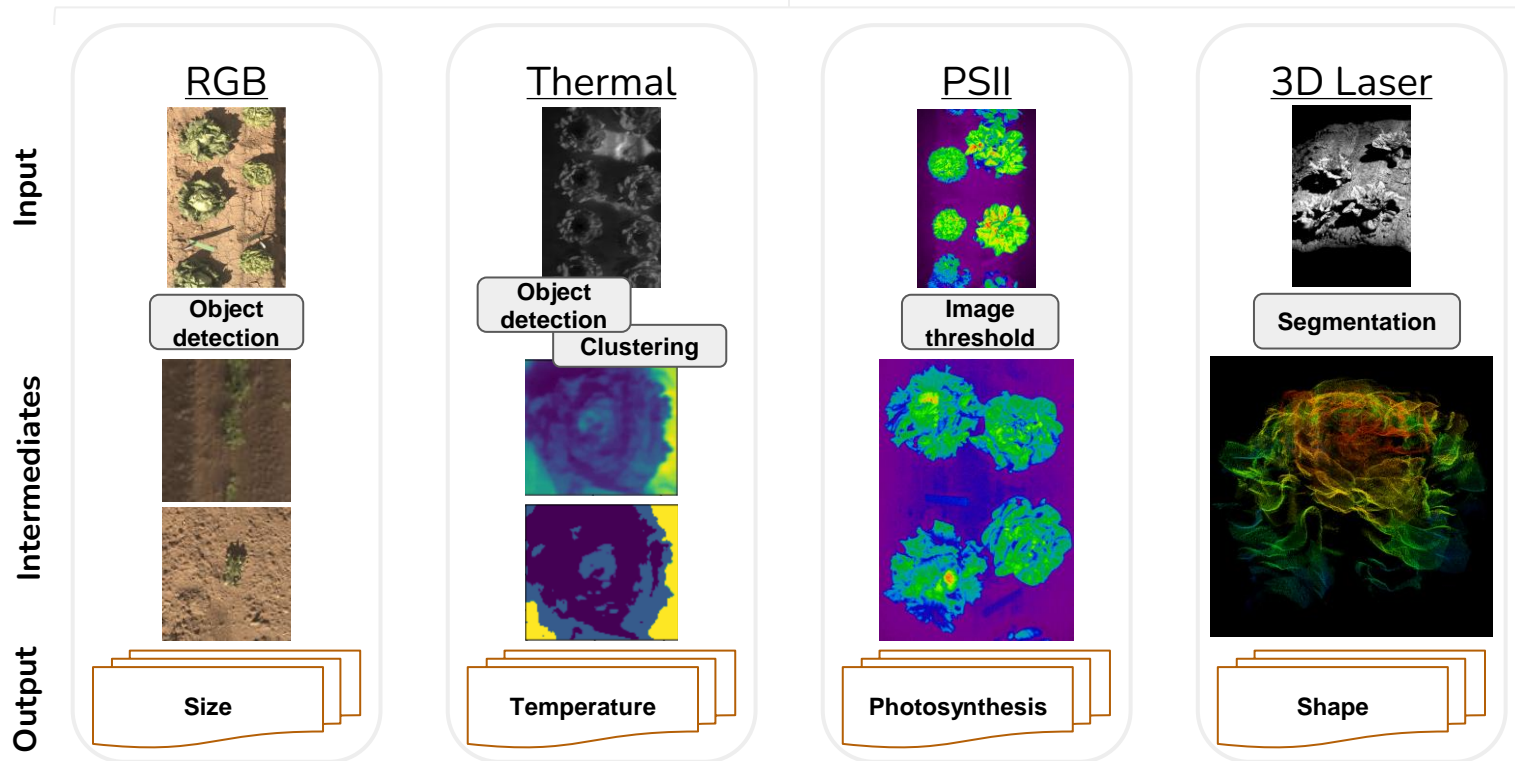
Thermal

20,000 plants * 36 scans * 4 phenotypes
= 2.9M individual plant phenotypes

Total: 6.2M time-series individual plant phenotypes



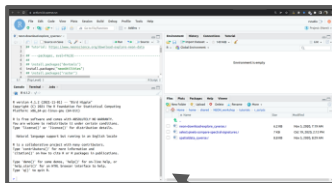
Requires modular & scalable workflows



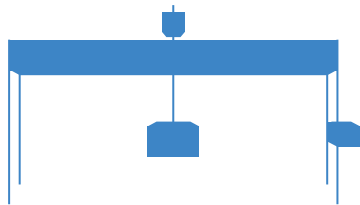
Data transfer and computation

Algorithms

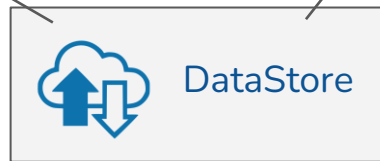
- Prototyping
- Development
- Testing
- Collaboration/Publication



Collaborators and Public



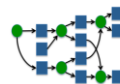
Cache Server



DataStore

iRODS

Makeflow



UA HPC

- Collect data
- 1-10TB/day

- Compression
- Checksums

- Storage
(raw + processed)

- HPC
- Workflows

Data transfer and computation

Figure 10.10.10

METHODS article

Front. Plant Sci., 06 March 2023

Sec. Technical Advances in Plant Science

Volume 14 - 2023 |




















<https://doi.org/10.3389/fpls.2023.1112973>

This article is part of the Research Topic

High-Throughput Field Phenotyping to Advance Precision
Agriculture and Enhance Genetic Gain, Volume II[View all 6 Articles >](#)

Download Article ▾

PhytoOracle: Scalable, modular phenomics data processing pipelines

 Emmanuel M. Gonzalez¹,  Ariyan Zarei²,  Nathaniel Hendler¹,  Travis Simmons¹,
 Arman Zarei³,  Jeffrey Demieville¹,  Robert Strand¹,  Bruno Rozzi¹,
 Sebastian Calleja¹,  Holly Ellingson⁴,  Michele Cosi^{1,5},  Sean Davey⁶,
 Dean O. Lavelle⁷,  Maria José Truco⁷,  Tyson L. Swetnam^{5,8},  Nirav Merchant^{4,5},
 Richard W. Michelmore^{7,9},  Eric Lyons^{1,4,5} and  Duke Pauli^{1,4*}

¹ School of Plant Sciences, University of Arizona, Tucson, AZ, United States² Department of Computer Science, University of Arizona, Tucson, AZ, United States³ Department of Computer Engineering, Sharif University of Technology, Tehran, Iran⁴ Data Science Institute, University of Arizona, Tucson, AZ, United States⁵ BIOS Institute, University of Arizona, Tucson, AZ, United States⁶ Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ, United States⁷ The Genome and Biomedical Sciences Facility, University of California, Davis, Davis, CA, United States⁸ School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, United States⁹ Department of Plant Sciences, University of California, Davis, Davis, CA, United States

23

total views

4

View Article Impact

Edited by



Sindhuja Sankaran

Washington State University, United States

Reviewed by



Nelson Nazzicari

Council for Agricultural and
Economics Research (CREA), Italy

Max Feldman


Agricultural Research Service, United States
Department of Agriculture,
United States

CyVerse Use Case # 2



Alex Bucksch

<https://www.computational-plant-science.org>

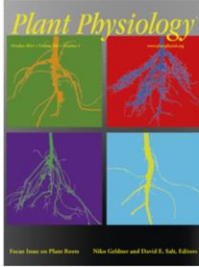
**Digital Imaging of Root Traits**
Getting to the roots of the crops!

[Login](#) | [Register](#) | [Help](#) | [Contact us](#)

HOME ABOUT GET STARTED ROOTS COMPUTATION

Digital Imaging of Root Traits

Digital imaging of root traits (DIRT) measures traits of monocot and dicot roots from digital images. DIRT automates the extraction of root traits by making high-throughput grid computing environment available to end-users without technical training.



Plant Physiology

Roots from the Plant Roots

Niko Colhofer and David E. Salt, Editors

DIRT measurements are inspired by the **Shovelomics** standard for root excavation. Users can compute over 70 phenotypic traits from images taken with the DIRT imaging protocol. To date, monocot and dicot roots along with excised root samples are possible to analyze. DIRT facilitates the use of high-performance computing systems as well as storage, organization and sharing of the image data. Our approach was highlighted on the **Plant Physiology** cover in October 2014. The software platform was published in **PlantMethods** in November 2015.

Unique features are:

- Join an active **googlegroup** to get help from users and developers
- Calculate root traits from large data sets (> 1000 images) imaged with the DIRT protocol
- Perform virtual experiments through recombination of existing experiments
- Store, share and organize images with in the whole user community, private or selected collaborators
- Retrieve calculations as excel compatible file or **RSML**
- Extend DIRT with python through open source (**Source Link**)
- Visual and statistical result control of all processing steps

Das, A., Schneider, H., Burrridge, J. et al. Digital imaging of root traits 2015 (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics. *Plant Methods* 11, 51 <https://doi.org/10.1186/s13007-015-0093-3>

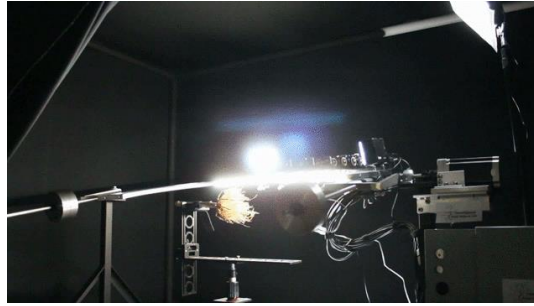


The system was financed in part by a seed grant from the Center for Data Analytics, Georgia Institute of Technology, Spatial Networks in Biology: Organizing and Analyzing the Structure of Distributed Biological Systems (A.Bucksch and J.S. Weitz), the NSF Plant Genome Research Program, NSF 0820624 (J.P. Lynch and J.S. Weitz) and an iPlant seed grant "High-throughput computing platform for quantifying root traits from images".

DIRT/3D measures root systems from the field



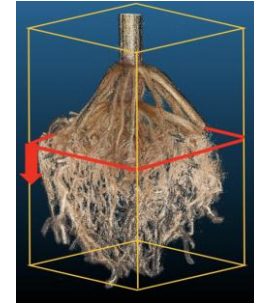
I. Maize roots in the field



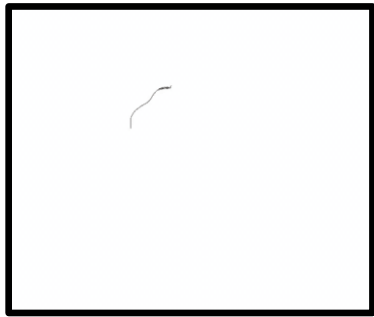
II. 3D root scanner



III. Root images data



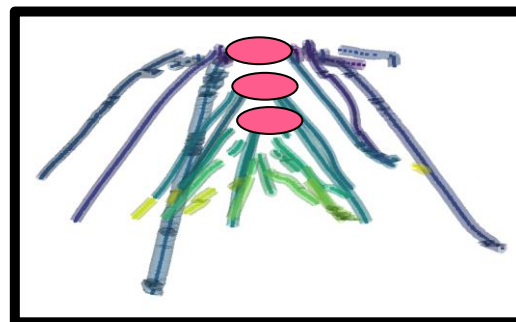
IV. 3D root model



V. Individual root tracking



VI. Computed root structure



VII. Resolved occluded whorl locations



Suxing Liu

DIRT/3D measures root systems from the field



Dense data sets are giant data sets

1 root = 1 GiB

1 experiment = 1,000 roots

Liu, S. et al. 2021 DIRT/3D: 3D root phenotyping for field-grown maize mays. Plant Physiology 187:2 <https://doi.org/10.1093/plphys/kiab311>

Home Workflows

- Projects
- Datasets
- Workflows
- Agents
- Tasks

Public Workflows

Yours powered by GitHub

Example: Working With Data

cyverse cyverse-discovery-environment data example

Computational-Plant-Science/plantit-example-accessing-data

A simple workflow mapping an input file to an output file.



DIRT3D Traits

phenomics phenotyping phenotyping-algorithms root

Computational-Plant-Science/3D_model_traits_demo

Computation of root phenes from 3D point clouds.


DIRT

computer-vis image-processing imaging phenotyping

phenotyping-algorithms root

Computational-Plant-Science/DIRT

Digital Imaging of Root Traits: Extract trait measurements from images of monocot and dicot roots.



Smart Plant Growth: Top-Down Traits From Images

computer-vision extract-traits image-segmentation phenotyping

Computational-Plant-Science/spg

Extract geometric traits from top-view images of plants.



Example: Hello World

bash example hello-world

Computational-Plant-Science/plantit-example-hello-world

Just about the simplest PlantIT workflow possible.



Herbarium Sheets

w-bonelli/herbarium-sheets

Image-based trait analysis for digitized herbarium sheets




Example: Parameters

example parameters

Computational-Plant-Science/plantit-example-parameters

A simple PlantIT workflow to demonstrate parameters.




Tomato Analyzer Lite

computer-vision fruit image-processing phenotyping segmentation

tomatoes traits

van-der-knaap-lab/tomato-analyzer-lite


Extract geometric traits from images of tomatoes, peppers, and more.



Stomata Counter

SvenTwo/epidermal

DCN stomata prediction on epidermal images




DIRT3D: Reconstruction

phenotyping phenotyping-algorithms root

Computational-Plant-Science/3D_model_reconstruction_demo

Implementation of the Visual Structure from Motion algorithm optimized for plant branching structures.




Vessel Detector

computer-vision microscopy phenotyping segmentation stems

w-bonelli/vessel-detector

Detect vessel elements in cross-sectional microscopy images of stem tissues.



<https://plantit.cyverse.org/>



THE UNIVERSITY
OF ARIZONA

Three PostDoc positions in root phenomics

Algorithms - Experiments - Plants

Apply: www.bit.ly/RootPhenotyping

**Discover unknown
phenotypes with computers**

**Address climate change in
drought & stress environments**

**Share open-source tools
with the plant community**

**Characterize function in
real agricultural settings**

**You have worked with
plants and trait data**

**You know python, imaging,
and experimental design**



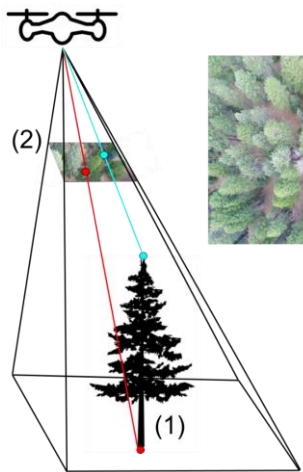
**Shape the future of plant phenotyping
& receive benefits and \$60,000+
(dependent on experience)**

**More info: www.computational-plant-science.org
Questions: bucksch@arizona.edu**



CyVerse Use Case # 3

Open Forest Observatory



Michael J. Koontz
mikoontz@gmail.com



Derek J. Young
djyoung@ucdavis.edu



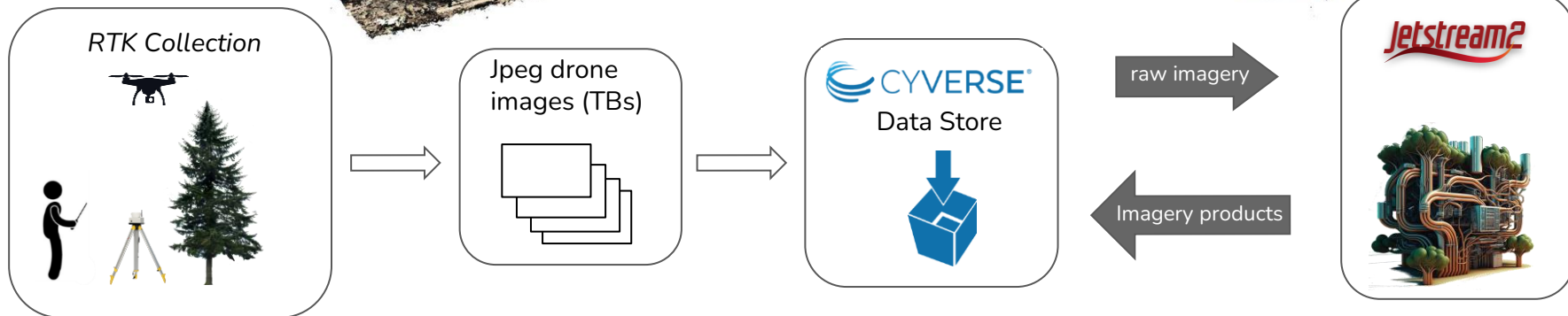
[Award # DBI 2152671](#)
[Award # DBI 2152672](#)
[Award # DBI 2152673](#)



Jeffrey K. Gillan
jgillan@arizona.edu



Collection Pipeline



COG
CLOUD OPTIMIZED
GEOTIFF



COPC.io
Cloud Optimized Point Cloud



Art by MidJourney AI



Analysis Pipeline

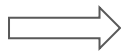
Jetstream2



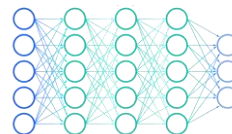
Photogrammetry

Worker VMs

Server VM



Individual Tree Detection
ML Species Classification



Fir
Pine
Spruce

<https://github.com/open-forest-observatory/automate-metashape>



Cloud Optimized Formats

Vectors



geojson.io



Rasters

STAC

SpatioTemporal
Asset Catalog



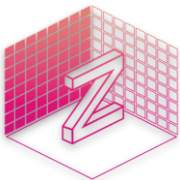
COG

CLOUD OPTIMIZED
GEOTIFF

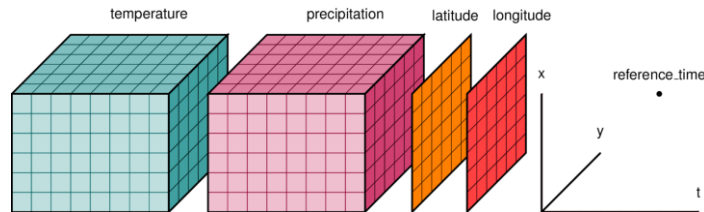


COPC.io

Cloud Optimized Point Cloud



Zarr



Conclusion

The Big Data Landscape -  not an exaggeration

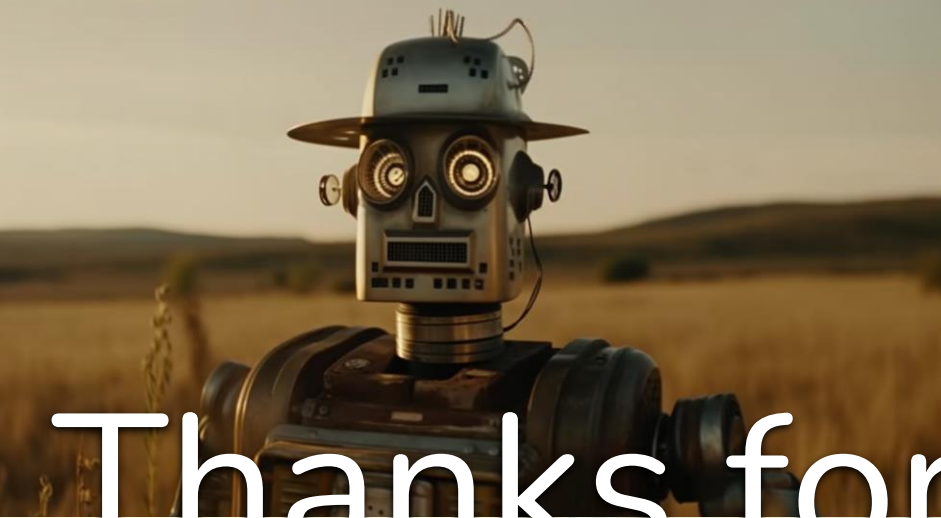
2023 Year of Open Science -  12/2025 

Becoming Cloud-Native -  not an option

AI in the workplace -  already changing the world



DBI-0735191, DBI-1265383, DBI-1743442 and OAC-1664172



Thanks for your time!