SGCI

Connecting people and resources to accelerate discovery by empowering the science gateway community

Summer Coding Institute, Part 2: Science Gateways

Marlon Pierce

Indiana University

https://sciencegateways.org

Award Number
ACI-1547611

NSF

What have you learned so far?

Learning the Stack, Part 1


Unix and Linux


Python


Software version control with Git


Jupyter notebooks


Using high performance computing

SGCI

Learning the Stack, Part 2

- Learn what science gateways can do
- Get tutorials on specific science gateway software stacks
  - Tapis
  - HUBzero
  - Apache Airavata
- And then build your own gateway prototype
  - COVID-19 data
  - Put your Python, Web, Git, Jetstream, and Docker knowledge to work

SGCI

SGCI

Connecting people and resources to accelerate discovery by empowering the science gateway community

Science Gateways: An Overview

Marlon Pierce

Indiana University &
Science Gateways Community Institute (SGCI)

https://sciencegateways.org

Award Number
ACI-1547611

NSF

What is a science gateway?

What is a gateway?

- An online community space for science and engineering research and education
- A Web-based resource for accessing data, software, computing services, and scientific instruments
- A way to help researchers, educators, students, and the public gain access to sophisticated or limited resources

SGCI

dREG: discriminative Regulatory Element detection from GRO-seq

- Prof. Charles Danko and Dr. Zhong Wang, Cornell
- dREG: Identification of the genomic regions that regulate transcription
  - DNA -> RNA -> Proteins
  - Genetic basis of many diseases

Problem: What's the best way to deliver the code?

- You want other people to use it
- It's actively developed
- It really needs a large GPU cluster to run well.

dREG: Software as a Service

Instead of requiring users to download and install the code, deliver it through a gateway.

And integrate third party Web tools

dREG
Discriminative Regulatory Element Detection

Home About Documents FAQ

Project Start dREG Saved dREG runs Storage Admin Dashboard DregEroma2017

Experiment Summary

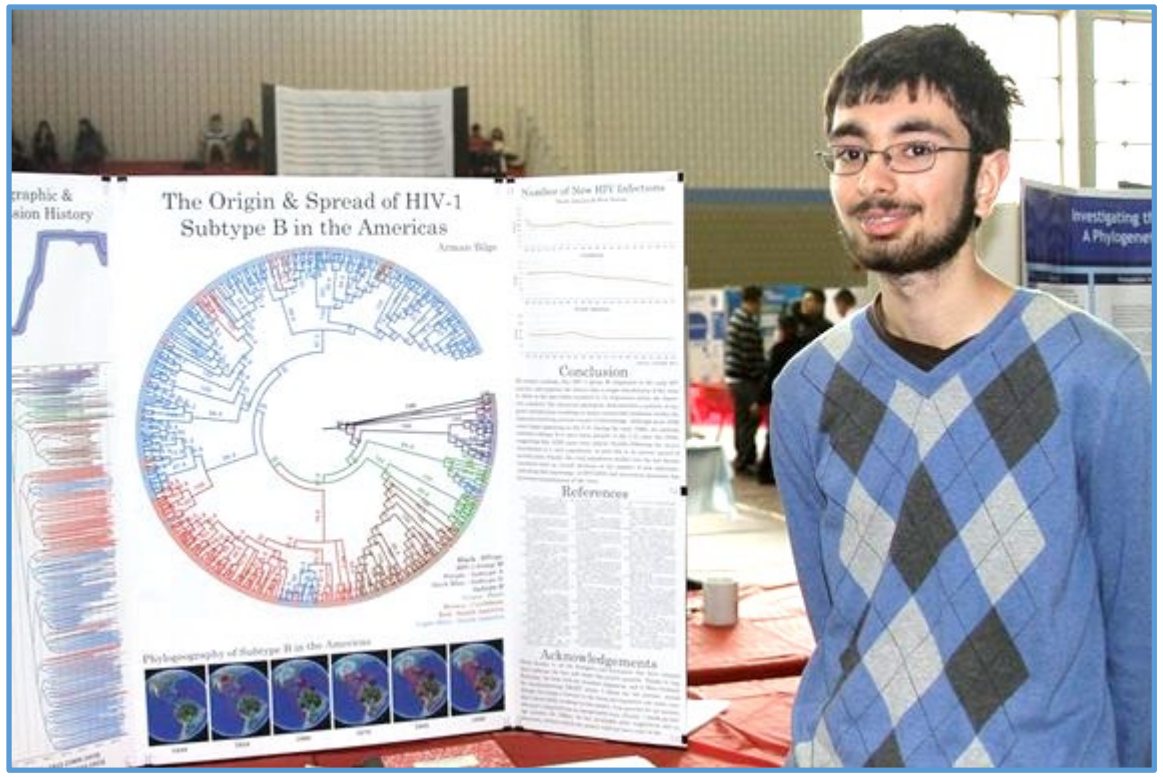
Enable Auto Refresh ON OFF

Experiment ID	dREG-Run1_cf6f1c3b-a79e-4a0e-bdfc-3bf389223cde
Name	dREG-Run1
Description	
Project	*You don't have access to this project.*
Owner	dregeroma2017
Application	dREG prediction
Compute Resource	comet.sdsc.edu
Experiment Status	COMPLETED
Job	Name: A285457678 · ID: 10026175 · Status: COMPLETE · Creation Time: 07/06/2017, 7:25 AM - GMT-0400 (EDT)
Notifications To:	eroma.abeysinghe@gmail.com
Creation Time	07/06/2017, 7:25 AM - GMT-0400 (EDT)
Inputs	BigWig(+strand): K562.chr21.plus.bw BigWig(-strand): K562.chr21.minus.bw
Outputs	Select results Download
Genome Browser	hg19 or input Switch to genome browser
Storage Directory	Open
Errors	

DANKO LAB · BAKER INSTITUTE FOR ANIMAL HEALTH · Cornell University

follow us on:

CIPRES: Cyberinfrastructure for Phylogenetic Research



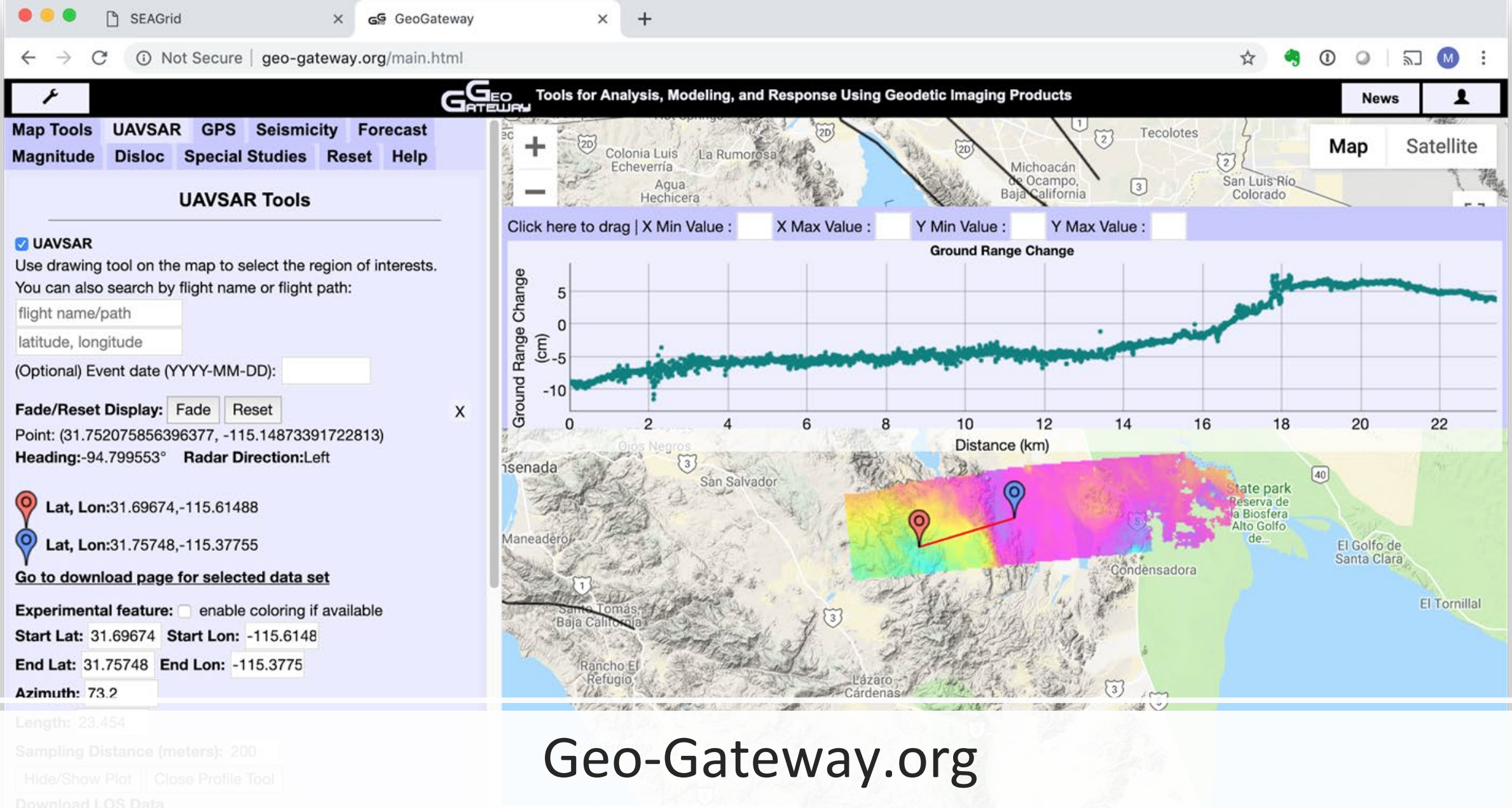
Armand Bilge, 10th grader at Lexington High School, next to a poster explaining his award-winning research project, a map and timeline that identified when HIV arrived in the Americas.

CIPRES SCIENCE GATEWAY

CIPRES | Home | Toolkit | Help | How to Cite Us

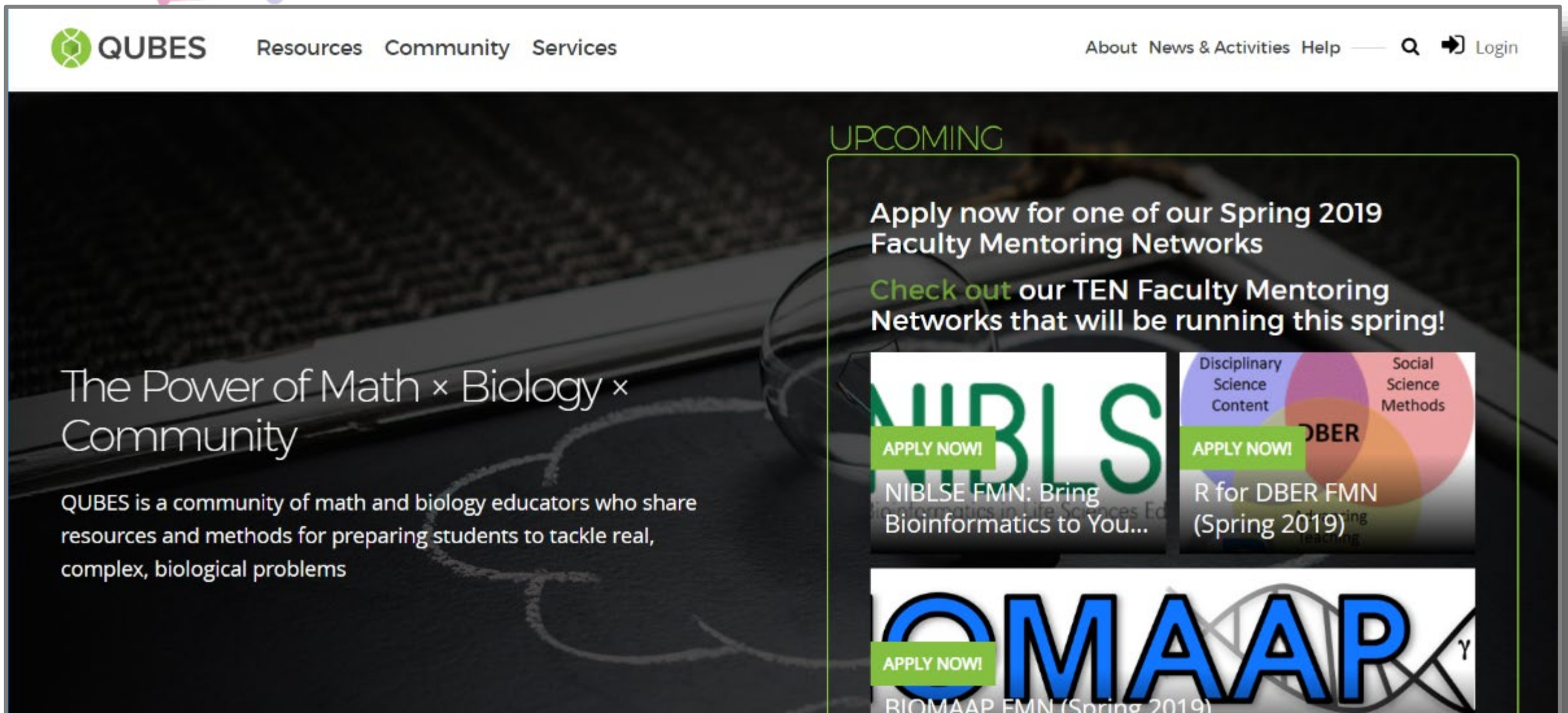
- **1.36 million jobs** on TeraGrid/XSEDE submitted by **33,195 unique users**.
- Used for **curriculum delivery** by at least 93 instructors.
- Supported **4,500+ publications**.

https://www.phylo.org/portal2



Geo-Gateway.org

QUBES: Quantitative Biology Education Resources



Citizen Science: Zooniverse

WHAT WILL YOU DISCOVER?

Participate in research of all kinds, from classifying galaxies to counting penguins to transcribing manuscripts. Whatever your interest, there's a Zooniverse project for you.

GALAXY ZOO

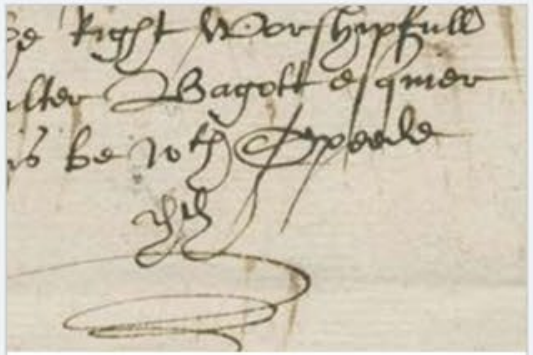
Help us discover the secrets of galaxy evolution by classifying distant galaxies.

View Project

CHIMP & SEE

Discover the secret life of chimpanzees. We need your help to study, explore, and learn from…

View Project



SHAKESPEARE'S WORLD

Transcribe handwritten documents by Shakespeare's contemporaries and help us understand his life an…

View Project

MUON HUNTER

Help astronomers to find elusive muons disguised as gamma rays!

View Project

See All Projects

SGCI

So what have we learned about science gateways?

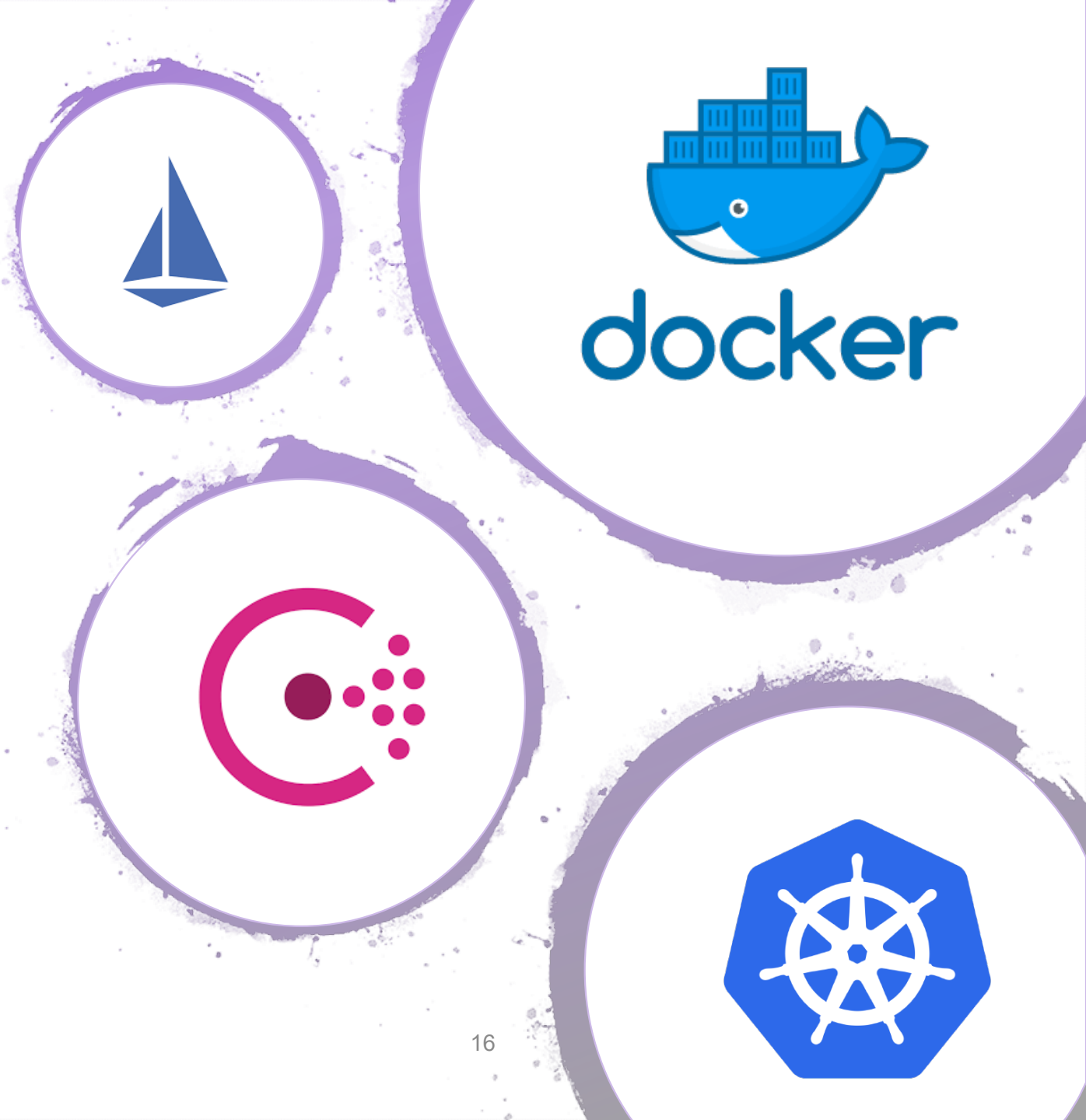
SGCI

#1 Gateways Are Like Startups

- Most of these gateways were built by small university teams
- Someone has an idea and wants to provide a service to a research and education community.
- Many fail but a few thrive.
- Thriving gateways still need help
  - Scale up
  - Sustain
  - Evolve

#2 Science Gateway Software Should Use Cloud Native Approaches

- Science gateway software systems are examples of distributed systems, cloud-native technologies



#3 Use Science Gateways to Teach Distributed Systems

- See https://courses.airavata.org
- Teach open source practices with GitHub
- Teach modern software engineering
  - Continuous Integration and Deployment
  - DevOps
  - Infrastructure as Code
- Build stuff! This is how you learn

GitHub

SGCI

Some Common Characteristics of Science Gateways

Patterns and abstractions

Pattern #1: Authentication

- You have to log in!
- Gateways authenticate their users
- Why?
- OpenID Connect is an important standard
  - Lots of support

SGCI

Pattern #2: Sessions

- You can think of these as shopping carts
- A powerful feature of many gateways is the ability to store information for you on your usage history
  - Reproducibility for an online experiment

SGCI

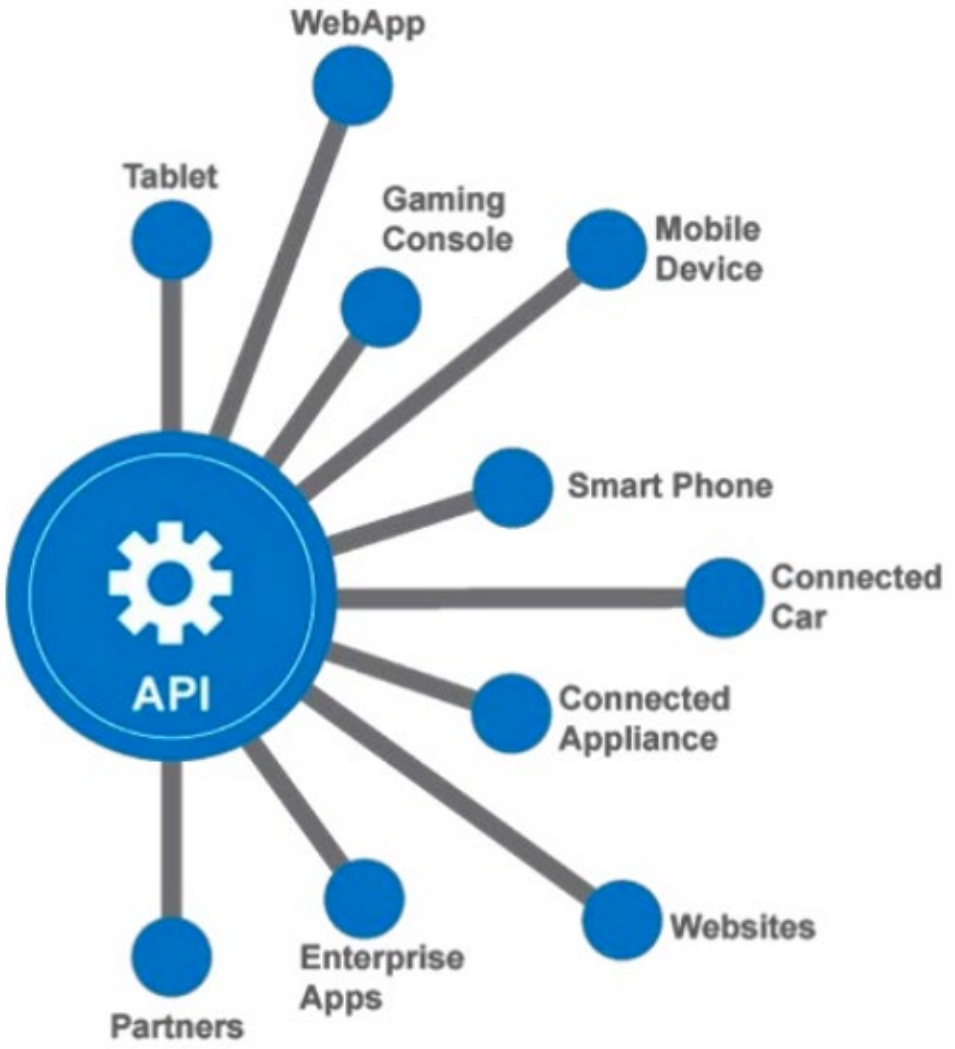
Pattern #3: Gateways Support Sharing and Publication

- Share your shopping carts!
- OK, this may be a bad idea for online shopping
- But it is a great idea for science
- Ultimate sharing in science is publication in a peer reviewed journal
- But before that, you may want to share with your advisor, colleagues, etc

SGCI

Pattern #4: APIs

- Many gateways can be accessed by different clients through Application Programming Interfaces
- And many gateways integrate 3rd party services through APIs.
- This is an architectural pattern more than capability pattern



SGCI

SGCI

Connecting people and resources to accelerate discovery by empowering the science gateway community

Outline for the Rest of the Institute

https://sciencegateways.org

Award Number
ACI-1547611

NSF

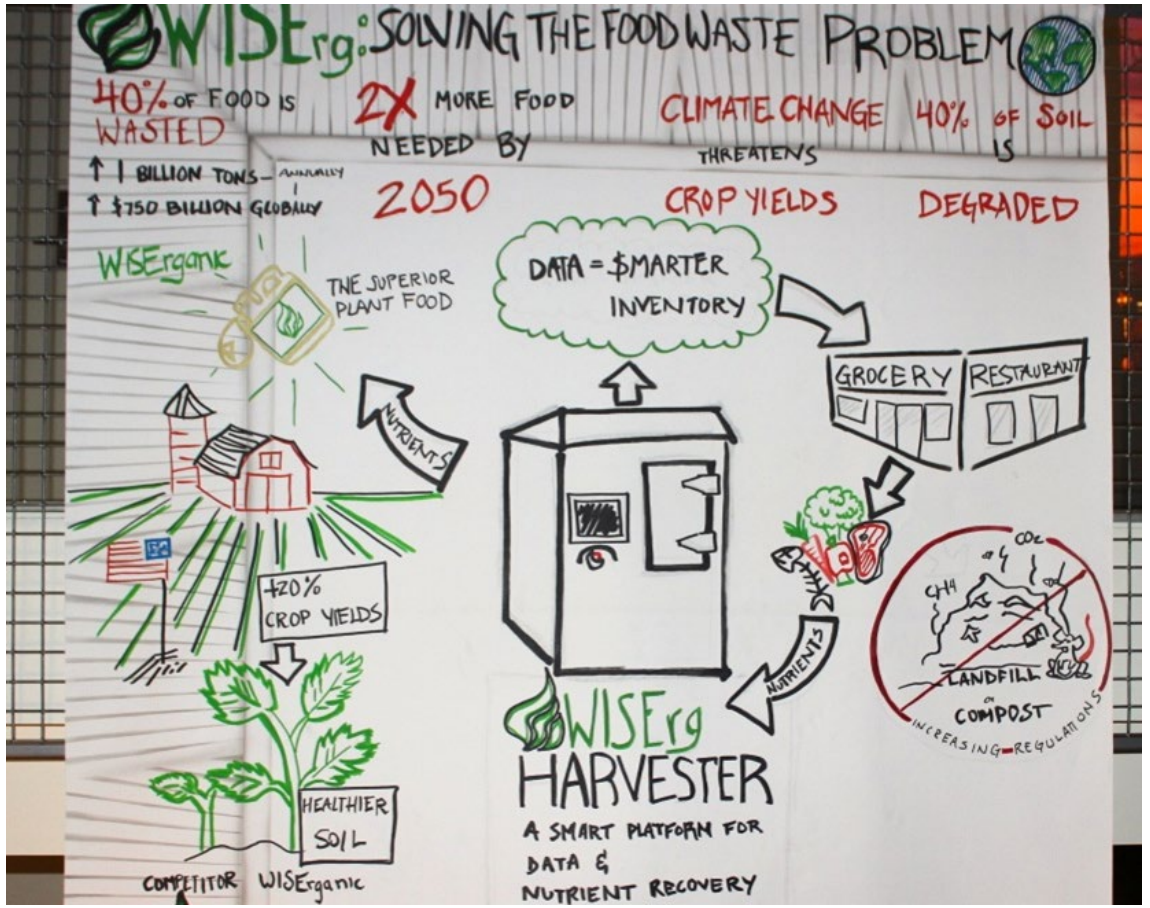
Outline for the Rest of the Institute

- You will divide into three teams
- Each team will build a gateway for accessing COVID-19 and related data.
- But first, you need to brainstorm

SGCI

July 16 -17: Napkin Drawing

- Each team will come up with a basic idea for a gateway
- Draw it in one figure
- Pitch it to everyone
  - July 17 afternoon session
- More in a moment



Team Implementations: July 20 -24

- Each team will implement their napkin drawing idea.
- You can use a simple tech stack (Python Flask, HTML, JavaScript)
- Or you can use one of the gateway frameworks from the tutorials

SGCI

July 22nd

- Prof. Paul Parsons will introduce usability and user experience for science gateways
- You can apply what he teaches to your own gateways

SGCI

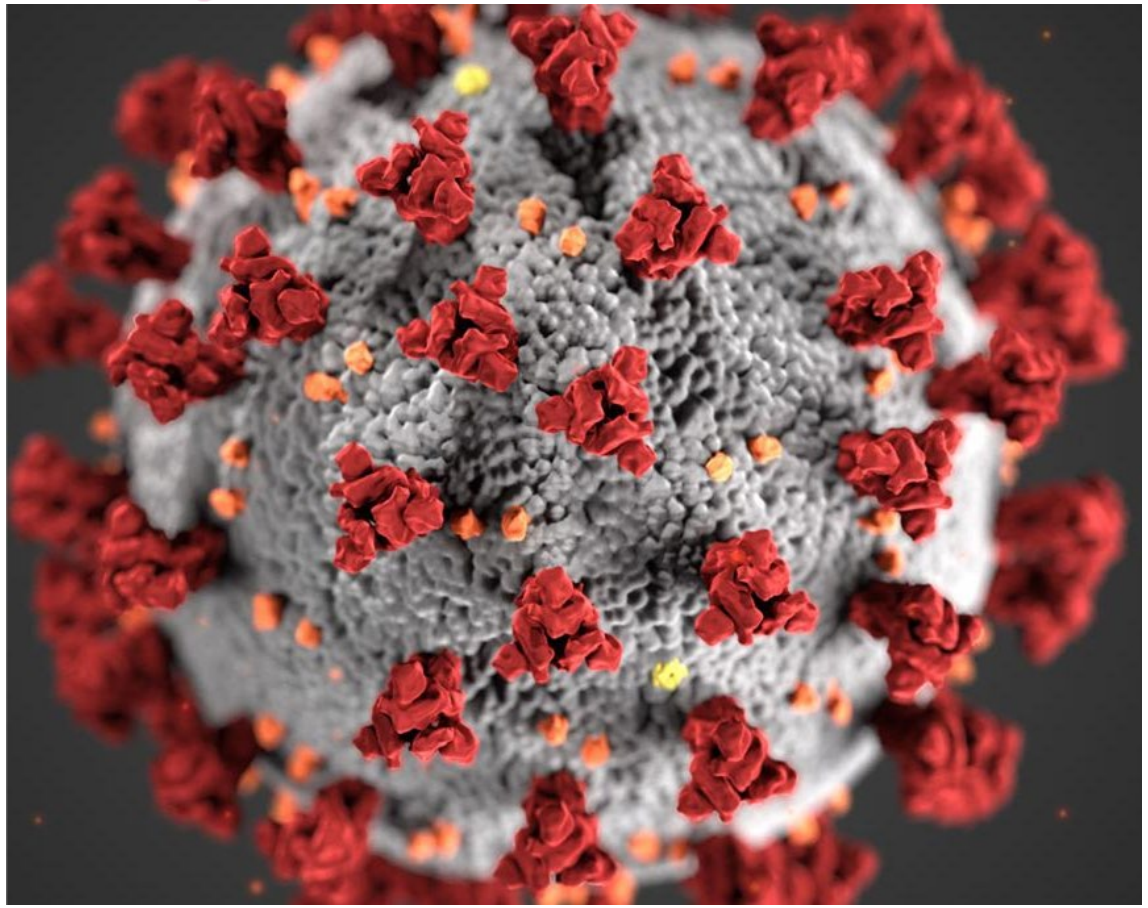
July 24: Final Presentations

- Each team will present and give a demo of your gateways

SGCI

Assignment 1: Project Brainstorming and Napkin Drawings

Your Team Projects



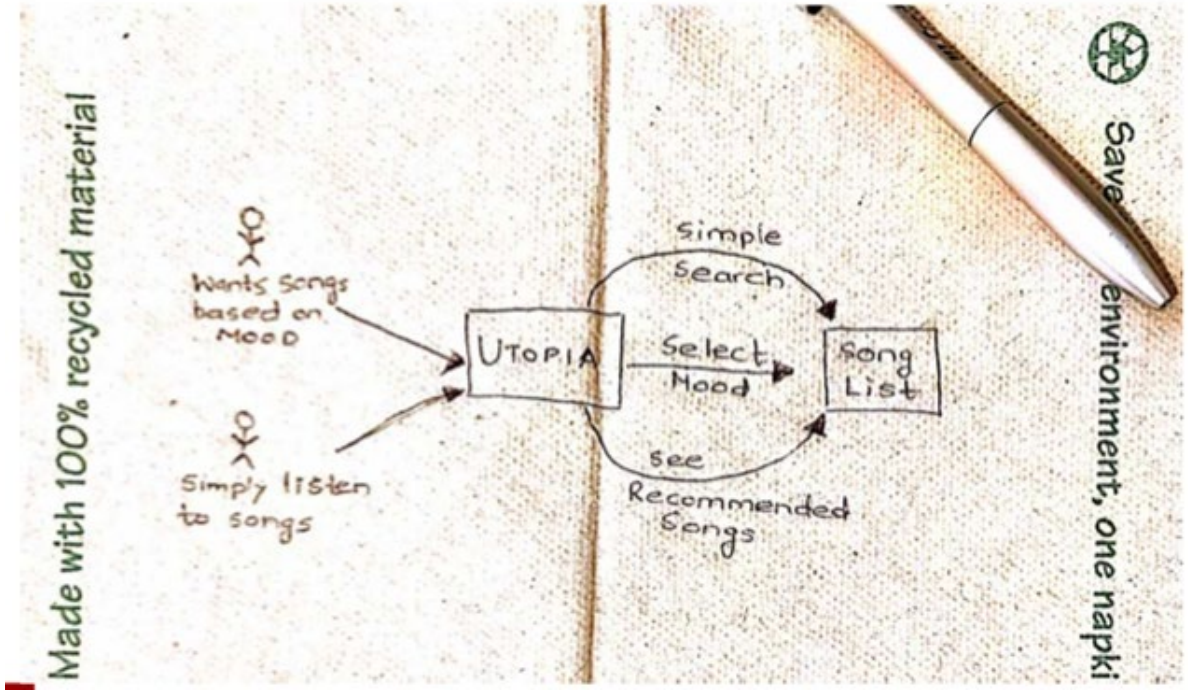
- Use today to brainstorm a prototype COVID-19 data gateway that you can build.
  - Review available data
  - Come up with at least 3 ideas
  - Pick one idea
  - You'll have mentors to help you
- You'll implement these ideas over the next several days

SGCI

Juliana Casavan on Napkin Drawings

- https://www.kaltura.com/index.php/extwidget/preview/partner_id/983291/uiconf_id/23620631/entry_id/0_uehdu0dv/embed/dynamic?&flashvars[streamerType]=auto

Good Napkin Sketches



All you need is one good slide

- What is your idea?
- Why would someone want to use it?
- How is your idea different?
- Avoid technical jargon and implementation details

SGCI

Schedule for July 16 -17

- July 16
  - Review available COVID-19 data
  - Brainstorm ideas
  - Come up with a team name for your GitHub repository
- July 17
  - Morning, Part 1: Technology overview for implementing gateways
  - Morning, Part 2: Prepare napkin drawing presentations
  - Afternoon: Each team gives their napkin drawing presentation “pitches”

SGCI

Starter COVID -19 Data Resources

- SafeGraph
  - https://docs.google.com/spreadsheets/d/1UNWvPzkUTTIXBZ6M6iGhM_7sr8h-MxsZdE7iOszkAmk/edit#gid=0
  - https://www.safegraph.com/covid-19-data-consortium
- Johns Hopkins
  - https://github.com/CSSEGISandData/COVID-19
- New York Times
  - https://github.com/nytimes/covid-19-data
- State data for your home states

SGCI

Some Questions for Brainstorming

- What do you want to know about COVID-19?
- Do you have any family or friends that are at high risk?
- Do you think we are taking this seriously enough? Too seriously?
- What is your school doing for Fall 2020 semester?

SGCI