

STATE AND FUTURE OF CI FOR LARGE FACILITIES

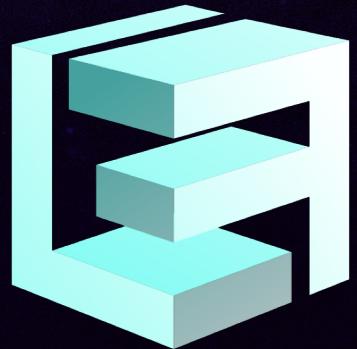
Stuart Anderson, Cal Tech/LIGO

Margaret Johnson, NCSA, LSST

Eric Lyons, U of A, Cyverse

Moderator: Dan Stanzione, TACC





LEADERSHIP-CLASS COMPUTING FACILITY



DISCUSSION QUESTIONS

- ▶ What has changed since 2017?
- ▶ What are the major challenges today and 5 years from now?
- ▶ What can be done at no additional cost?
- ▶ Where can collaborations and sharing help (hw, sw, people)?



Panel: State and Future of CI for Large Facilities

Margaret W. G. Johnson

2019 NSF Workshop on Connecting LFs and CI

September 16, 2019



NCSA | National Center for
Supercomputing Applications

Introduction

- Margaret Johnson
- Assistant Director at NCSA
- Involved in managing construction (MREFC) of the LSST Data Facility
- Involved in planning and implementing transition to LSST (pre-)operations
- Involved in discussions about planning CI for Multi-Messenger Astrophysics

Observations

- Facility projects take on building a lot of their own components.
- Naturally, staff hired onto projects think in terms of their components.
 - Particular areas of experience and expertise, and their trusted communities.
- Projects are then reliant on (captive to) this frame of reference to select and build components.
- In MREFC, a project brings on staff who think about the technologies that will build the components, but may lose sight of the business purpose.
 - i.e., how the facility will need to operate.

CI in the Facility Lifecycle

- By design, MREFCs are supposed to have a WBS and focus on delivering it.
 - Hard to change an existing EVMS project, unless there is programmatic support for change.
- When a project is conceptualized, the right balance of people is needed to
 - bring expertise from the CI community.
 - bring those who have a frame of reference oriented towards operations.
 - bring those who have a frame of reference oriented towards reuse of CI capabilities.
 - develop the complete, risk-reduced plan.
- When a facility is in operations, the program environment provides an opportunity for re-integrating with the broader CI community.
 - Focus is on operation and evolution, and ultimate goal of research mission.
 - Able to look to the CI community for evolution of capabilities and continued best practices.
 - Able to contribute experience shared practices, shared capabilities, expertise.

Progress by Having a Process

- The LF & CI community can make progress by working strategically within the existing facility lifecycle and the respective program environments.
- Consider what mechanisms are in place to ensure there is a healthy balance of staffing.
 - E.g., review at conceptualization
- Operations staff are the ultimate stakeholders of the facility CI development. Consider setting up distinct core operations earlier in the MREFC to ensure the broader focus on the ultimate mission of the facility during implementation.
 - Influence implementation decisions (i.e., agile development).
 - Support reuse of CI.
 - Incrementally transition to operations.
- The CI Center of Excellence has an important programmatic relationship with this process.



Transforming Science Through Data Driven Discovery

Eric Lyons
School of Plant Sciences
Biosystems Engineering
College of Agriculture and Life Sciences
University of Arizona



CyVerse

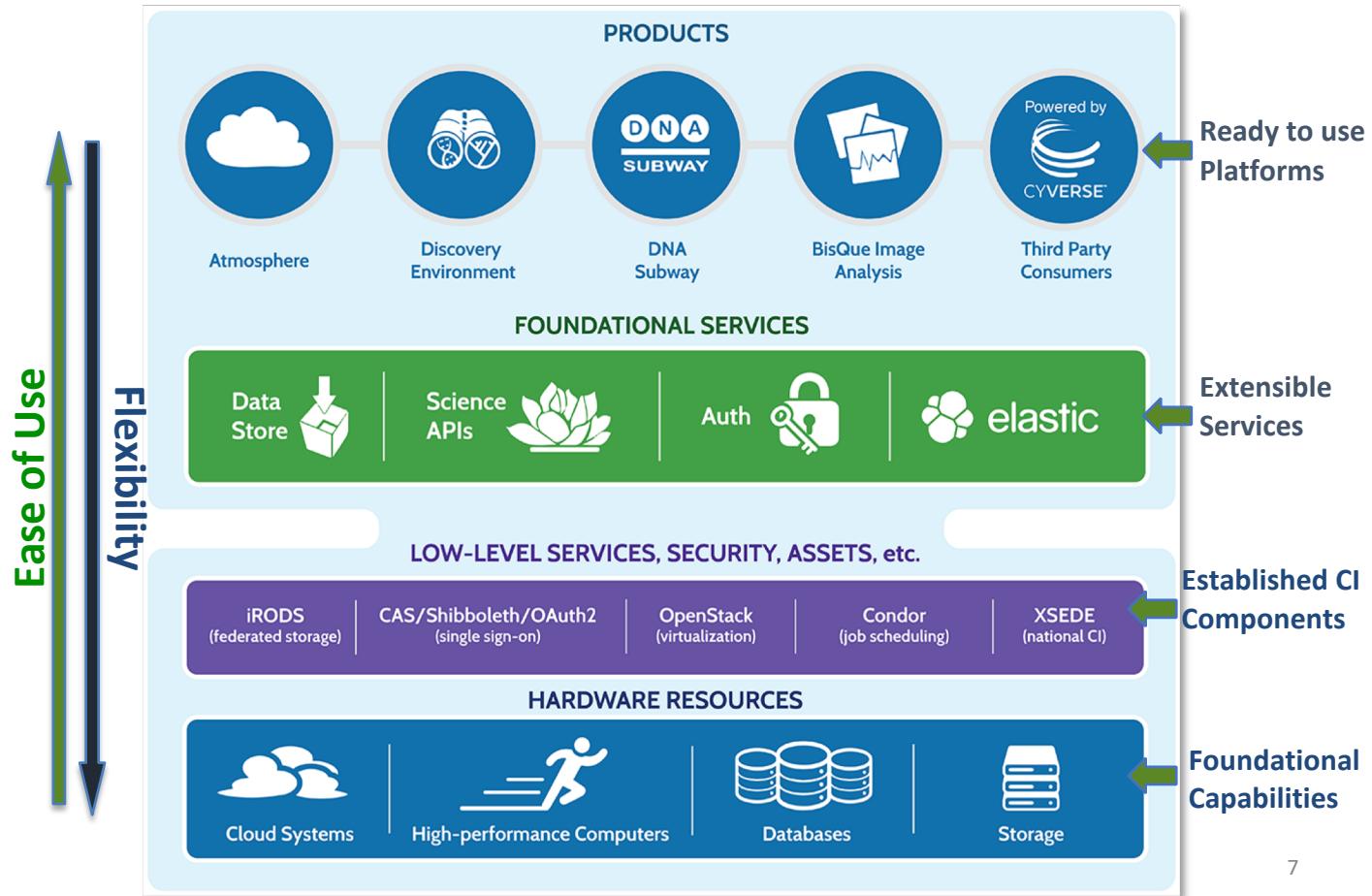
Mission: To design, deploy, and expand a national Cyberinfrastructure for Life Sciences research, and to train scientists in its use.

- \$115 Million, 15-year data science project
- Primary Funding from NSF BIO Directorate
- 65,000+ user accounts
- 130+ countries
- 7PB of data moves through CyVerse annually

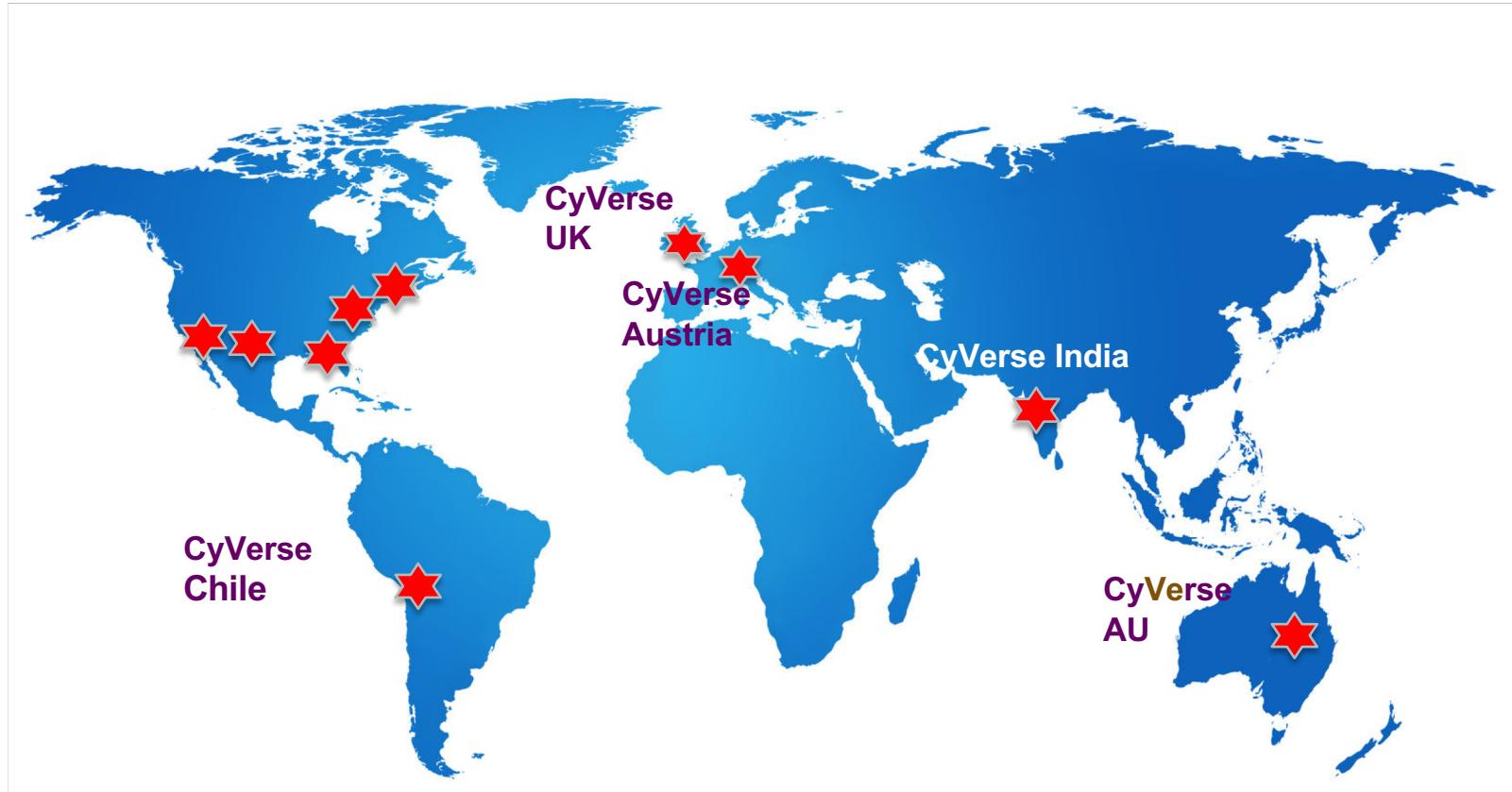


The CyVerse Stack

A Blueprint for Cyberinfrastructure Design



CyVerse: International Partners



CyVerse: Life Science and Beyond



Biomedicine



Plant Sciences



Animal Sciences



Geology
Climate



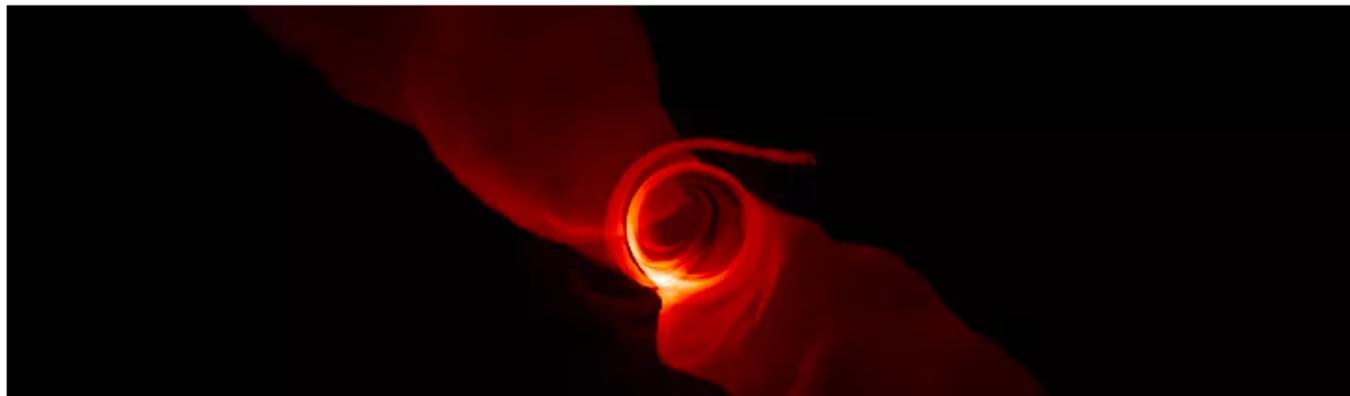
Space
Sciences



Ecology
Environment

CyVerse Community Member and Astrophysicist Helps Lead Event Horizon Telescope Data Processing

April 10, 2019 | Shelley Littin



A simulated image by the Chi-kwan Chan of University of Arizona shows the turbulent plasma in the extreme environment around a supermassive black hole. (Photo credit: University of Arizona.)

A CyVerse collaborator and astrophysicist helps lead data processing efforts for the [Event Horizon Telescope \(EHT\) project](#), an international collaboration aiming to capture the first image of a black hole by integrating eight telescopes to create a planet-sized virtual telescope capable of measuring the boundaries of black holes. The EHT revealed its [first black hole image](#) today at a [press conference](#) hosted by the [National Science Foundation](#).

CyVerse: What has changed since 2017

The image shows the CyVerse Discovery Environment interface with three main windows:

- Miles Per Gallon**: A browser window showing a scatter plot of Miles Per Gallon vs Cylinders.
- RStudio**: An RStudio window showing R code and a file tree.
- JupyterLab**: A JupyterLab window showing a Python notebook titled "Untitled.ipynb". The notebook contains code for reading a CSV file and creating a histogram. It also displays a taxonomic bar chart titled "qiime2_moving_picture.ipnb" with a color palette and a legend listing various bacterial phyla and genera.



ML, App Integration, Training, Scaling People

Transfer learning with VGG16 CNN



Upendra Kumar Devisetty
Instructor of [Big Data Fundamentals via PySpark](#)
7,500 STUDENTS
Science Analyst at CyVerse¹
[See Full Bio](#)



Ph.D in Genetics
The University of Nottingham

Science Informatician
CYVERSE™

- Computational biology and Bioinformatics
- Data science and programming instructor - Datacamp and Carpentries

1

2



CyVerse: Low cost synergies

Joint partnerships across programs
Training and workshops
Scaling people?

