



# CYVERSE™

Transforming Science Through Data-driven Discovery

## Containers & beyond at CyVerse

Ian McEwen  
(Software Engineer, University of Arizona)

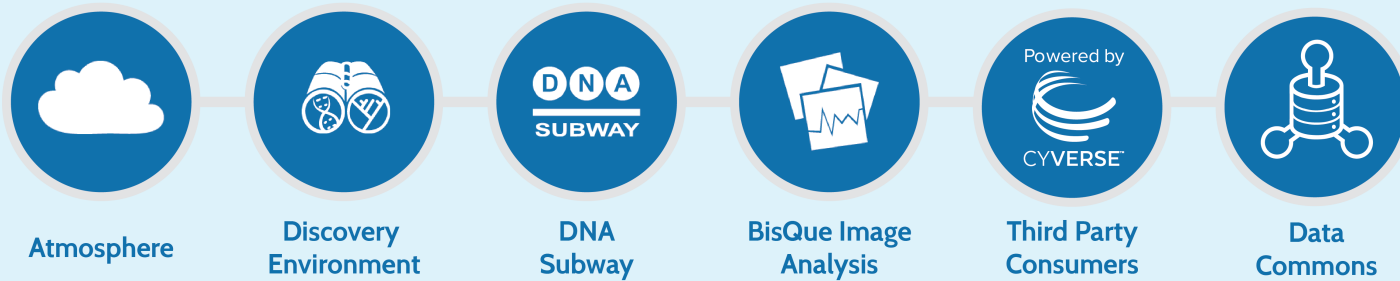


Cold  
Spring  
Harbor  
Laboratory



# CyVerse high-level architecture

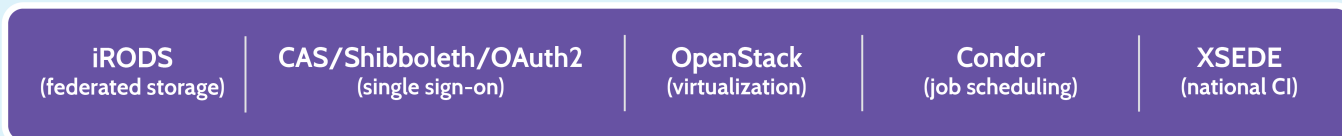
## PRODUCTS



## FOUNDATIONAL SERVICES



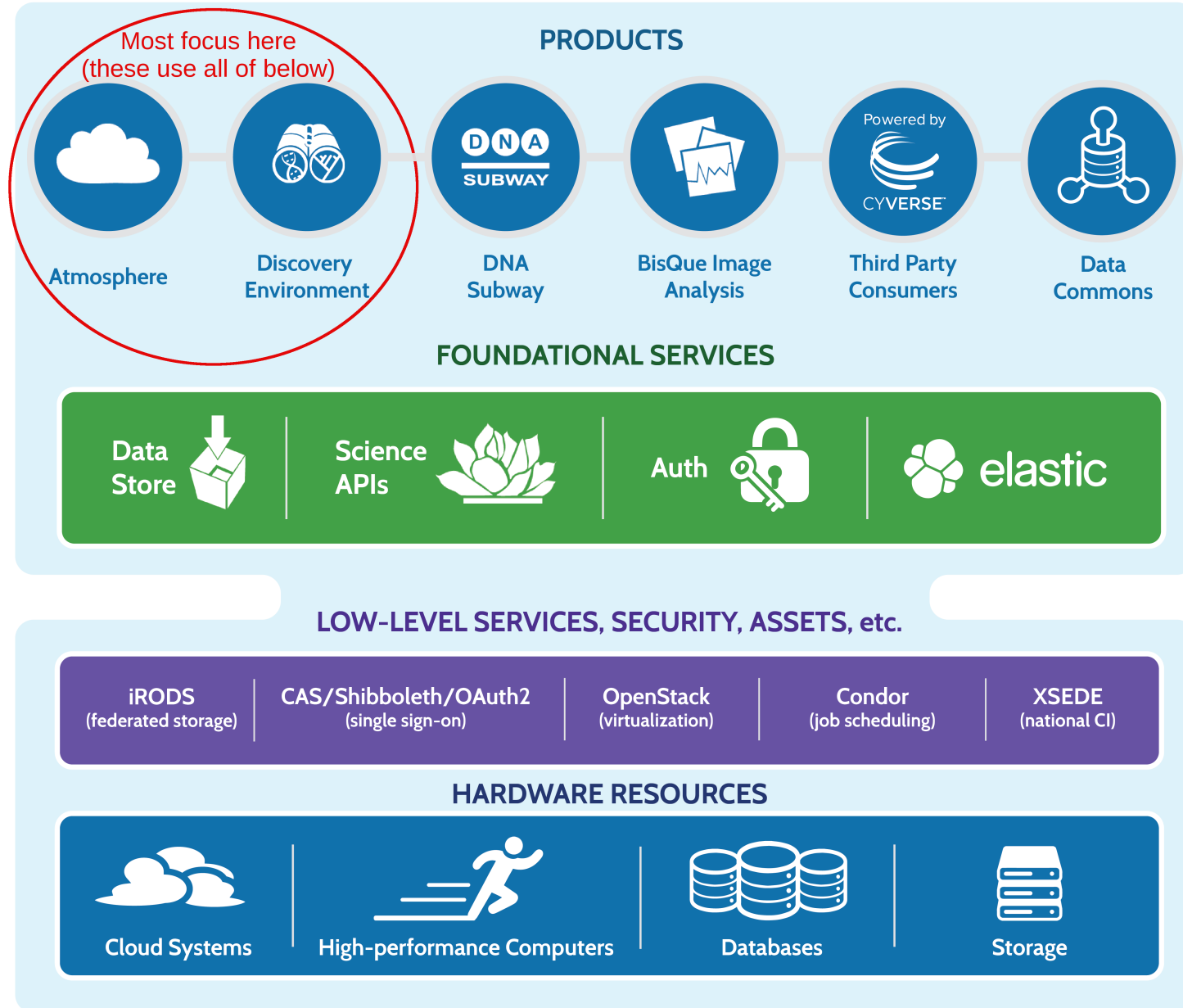
## LOW-LEVEL SERVICES, SECURITY, ASSETS, etc.



## HARDWARE RESOURCES



# CyVerse high-level architecture

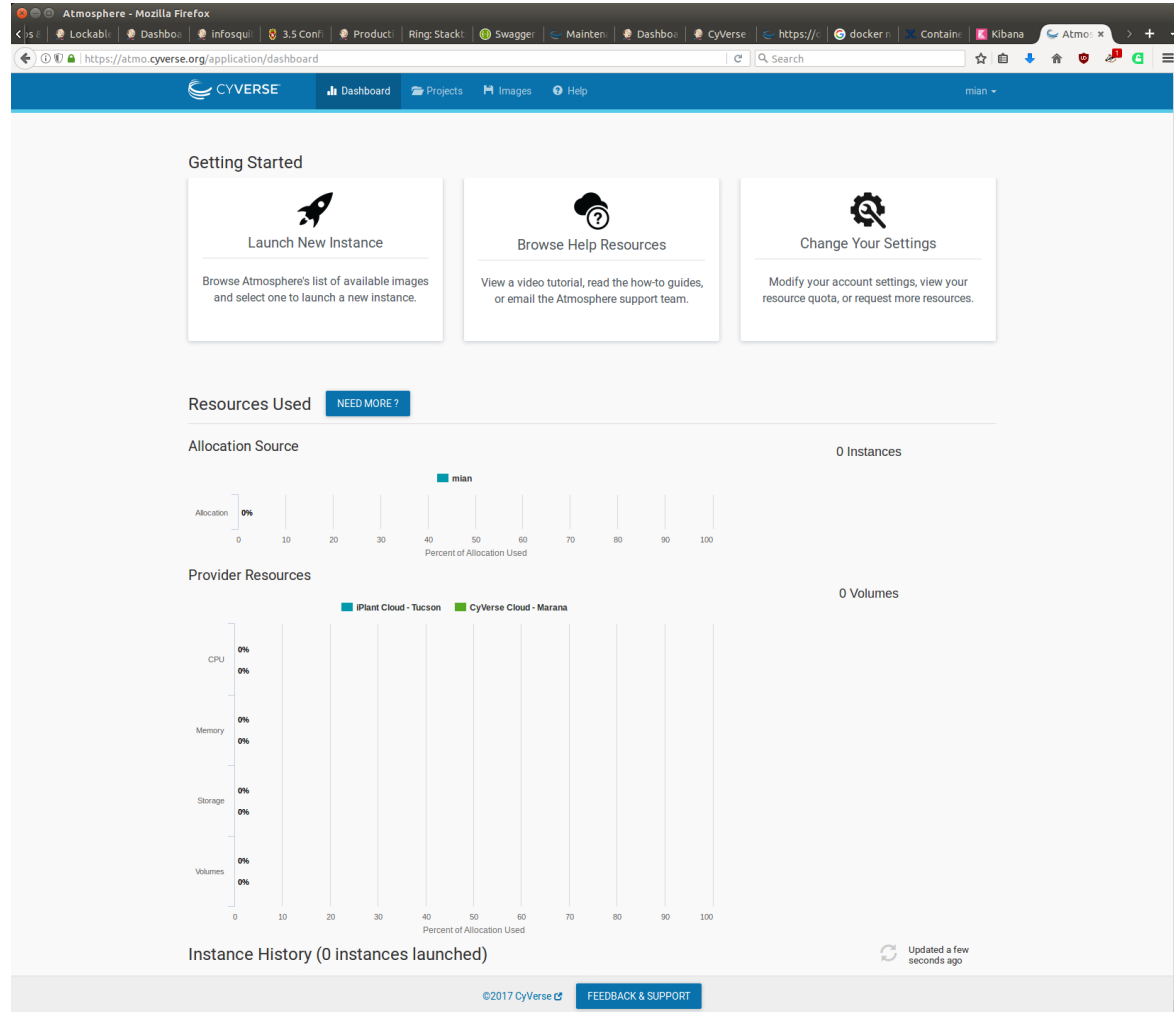


# Atmosphere

Atmosphere provisions VMs for use by users; catalog of precreated images to make it easy to get started.

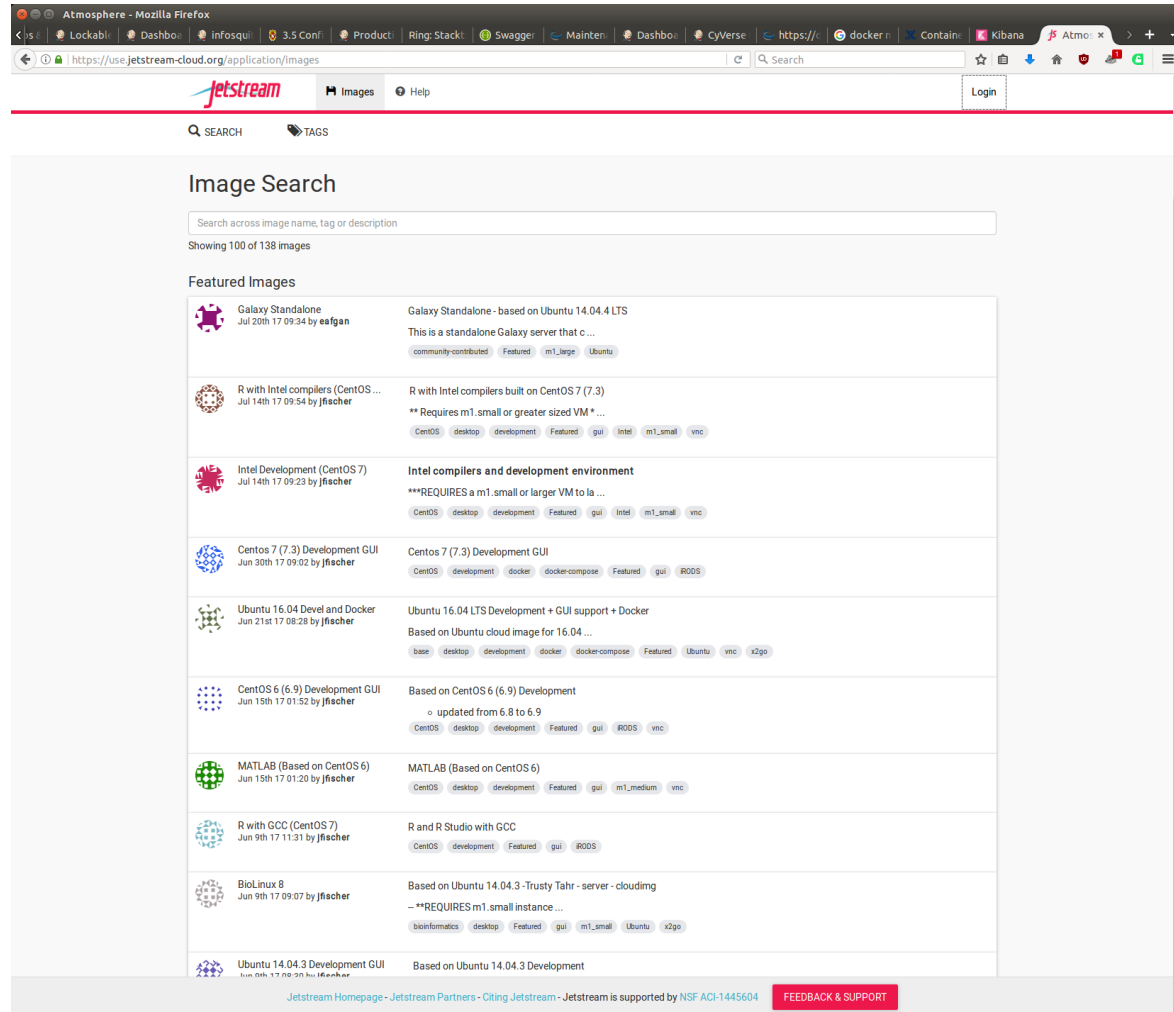
Atmosphere provides a place for technically savvy scientists and affiliates to experiment with technologies, e.g. containerization.

Data in the CyVerse Data Store can be accessed via an automatic FUSE mount and/or command line and GUI tools.



# Jetstream

Atmosphere, but using  
XSEDE login/others'  
(TACC/Indiana University)  
hardware.



The screenshot shows the Jetstream Image Search web application. The browser's address bar displays the URL <https://use.jetstream-cloud.org/application/images>. The page features a navigation bar with the Jetstream logo, a search bar, and a 'Login' button. Below the navigation bar, the 'Image Search' section includes a search input field and a message indicating 'Showing 100 of 138 images'. The 'Featured Images' section lists various pre-built image templates, each with a unique icon, title, creation date, creator, and a list of tags. The featured images include:

- Galaxy Standalone**: Based on Ubuntu 14.04.4 LTS. Tags: community-contributed, Featured, m1\_large, Ubuntu.
- R with Intel compilers (CentOS 7)**: Requires m1\_small or greater sized VM. Tags: CentOS, desktop, development, Featured, gui, Intel, m1\_small, vnc.
- Intel Development (CentOS 7)**: Requires a m1\_small or larger VM. Tags: CentOS, desktop, development, Featured, gui, Intel, m1\_small, vnc.
- Centos 7 (7.3) Development GUI**: Based on CentOS 7 (7.3) Development. Tags: CentOS, development, docker, docker-compose, Featured, gui, RTOOLS.
- Ubuntu 16.04 Devel and Docker**: Based on Ubuntu cloud image for 16.04. Tags: base, desktop, development, docker, docker-compose, Featured, Ubuntu, vnc, x2go.
- CentOS 6 (6.9) Development GUI**: Based on CentOS 6 (6.9) Development. Tags: CentOS, desktop, development, Featured, gui, RTOOLS, vnc.
- MATLAB (Based on CentOS 6)**: Based on CentOS 6. Tags: CentOS, desktop, development, Featured, gui, m1\_medium, vnc.
- R with GCC (CentOS 7)**: Based on CentOS 7. Tags: CentOS, development, Featured, gui, RTOOLS.
- BioLinux 8**: Based on Ubuntu 14.04.3-Trusty Tahr - server - cloudimg. Tags: bioinformatics, desktop, development, Featured, gui, m1\_small, Ubuntu, x2go.
- Ubuntu 14.04.3 Development GUI**: Based on Ubuntu 14.04.3 Development.

At the bottom of the page, there is a footer with the text 'Jetstream Homepage - Jetstream Partners - Citing Jetstream - Jetstream is supported by NSF ACI-1445604' and a 'FEEDBACK & SUPPORT' button.



# Discovery Environment (DE)

(the part I work on directly!)

The DE is the main workbench & entry-point into CyVerse infrastructure, aimed at less technical users and more established processes.

The screenshot displays the CyVerse Discovery Environment (DE) web interface. The top navigation bar includes links for Lockable, Dashboard, Info, Product, Ring, Stack, Swagger, Maintainer, Dashboard, CyVerse, and a search bar. The main interface is divided into several panels:

- Apps Panel:** Displays a grid of application cards, including CACE, DNANJ, DNAPARS, FastTree, Molecular Dating with Penalized Likelihood, and Ninja. Each card shows the application name, a brief description, and a star rating.
- Analyses Panel:** A table listing various analyses, including 'compress\_then\_wordcount\_analysis1', 'Word\_Count\_analysis1', and 'DE\_Word\_Count\_multi-file\_analysis1'. The table includes columns for Name, Owner, App, Start Date, End Date, and Status.
- Data: G-OnRamp\_hubs Panel:** A file browser view showing a directory structure of genomic datasets, including 'Bemisia\_tabaci', 'Centropomus\_pausiflorus', 'Drosophila\_miranda\_MS...', 'FraVesHawaii\_1.0', 'Ganaspis\_sp.1', 'JBrowse-1.12.3', 'JBrowse\_hubs', 'Red\_Siskin', 'Solenodon\_paradoxus', and 'Tetrahymena\_thermophila'.
- Collaboration Panel:** A window showing a list of collaborators, including 'Dennis Roberts', 'Srikanth Srinivasan', and 'Discovery Environment devs'.



# Managing lifecycles & pipelines

Since the DE aims to make the technical details largely transparent to the user, it broadly manages smaller and larger lifecycles and pipelines necessary to the process of doing computational work. For example:

- Data flows in, moves through various applications and analyses, management of data and metadata, eventually going on to be published with permanent identifiers in the CyVerse Data Commons or other repositories
- Apps/Analyses show a simple UI and job execution system handles staging and returning data, coordinating between systems (local vs. HPC), resource scheduling, permissions, notification, etc. This system is the main application of containerization in order to aid reproducibility, deployment, flexibility, etc.

Ongoing/future directions about expanding these ideas in many directions.



# Recent & Ongoing Work

- Private Tools – resources more heavily limited, but just paste in a docker image to use (*done, just released*)
- Groups & Teams – ongoing work, being integrated at low levels like resource allocation and data permissions as well as higher levels like a concept of a project, team, etc.
- Bring your own compute – enabling users to provide computational resources (perhaps via Jetstream) into the DE for their own use, bringing the DE and its features closer to them
- Interactive jobs – spin up e.g. Jupyter & RStudio images with dependencies of relevance to the science task inside the system which already manages permissions, data staging, notification, etc.
- Better and more flexible tools for high-throughput processing, resource allocation, etc.
- Integrating deeper with the broader community, especially common repositories like bioconda/biocontainers





Perpetually, much more to talk  
about than time!

(Questions?)

