



# CyVerse Overview

MCBIOS 2016 – University of Memphis, TN

**Jason Williams – Lead, CyVerse – Education, Outreach, Training**

Cold Spring Harbor Laboratory

williams@cshl.edu [ @JasonWilliamsNY]

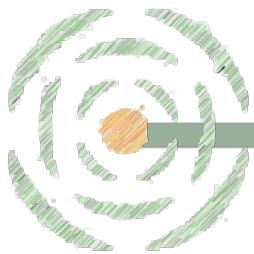


Download Slides and Follow Along

[mcbios.readthedocs.org](https://mcbios.readthedocs.org)



# CyVerse Evolution



***iPlant 2008***  
Empowering a New Plant  
Biology



***iPlant 2013***  
Cyberinfrastructure for Life  
Science



***CyVerse 2016***  
Transforming Science  
Through Data-Driven  
Discovery



# CyVerse Evolution



*DBI-0735191 and DBI-1265383*

We are funded by the National  
Science Foundation

- We are your colleagues and collaborators!
- \$100 Million in investment
- Freely available to the community
- Spur national/international collaboration
- Cite CyVerse:  
[CyVerse.org/acknowledge-cite-cyverse](http://CyVerse.org/acknowledge-cite-cyverse)



# CyVerse Evolution



***CyVerse 2016***

Transforming Science  
Through Data-Driven  
Discovery



## Vision:

Transforming science through data-driven discovery

## Mission:

Design, develop, deploy, and expand a national  
**cyberinfrastructure** for life science research, and train  
scientists in its use

**More than 30K users, PB of data, and hundreds of  
publications, courses, and discoveries**



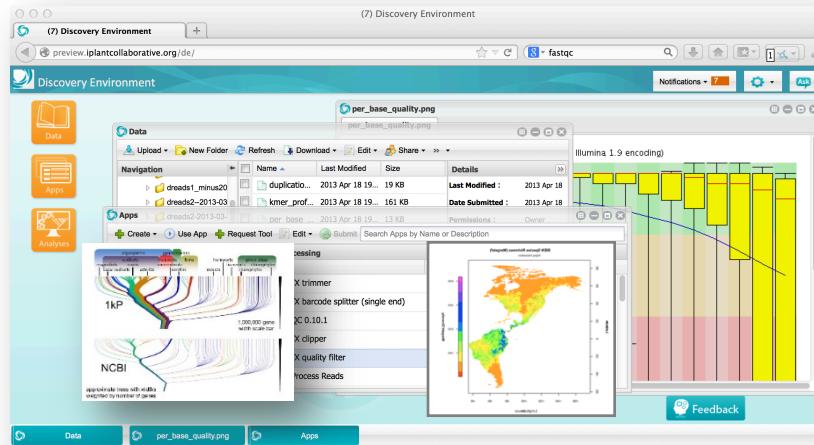
# What is Cyberinfrastructure?

- Data storage
- Software
- High-performance computing
- People

organized into systems that solve problems of size and scope that would not otherwise be solvable.



# What is Cyberinfrastructure?



Platforms, tools, datasets



Storage and compute



Training and support

# CyVerse supports all domains of life science



Plant / Microbial



Animal



Biomedical

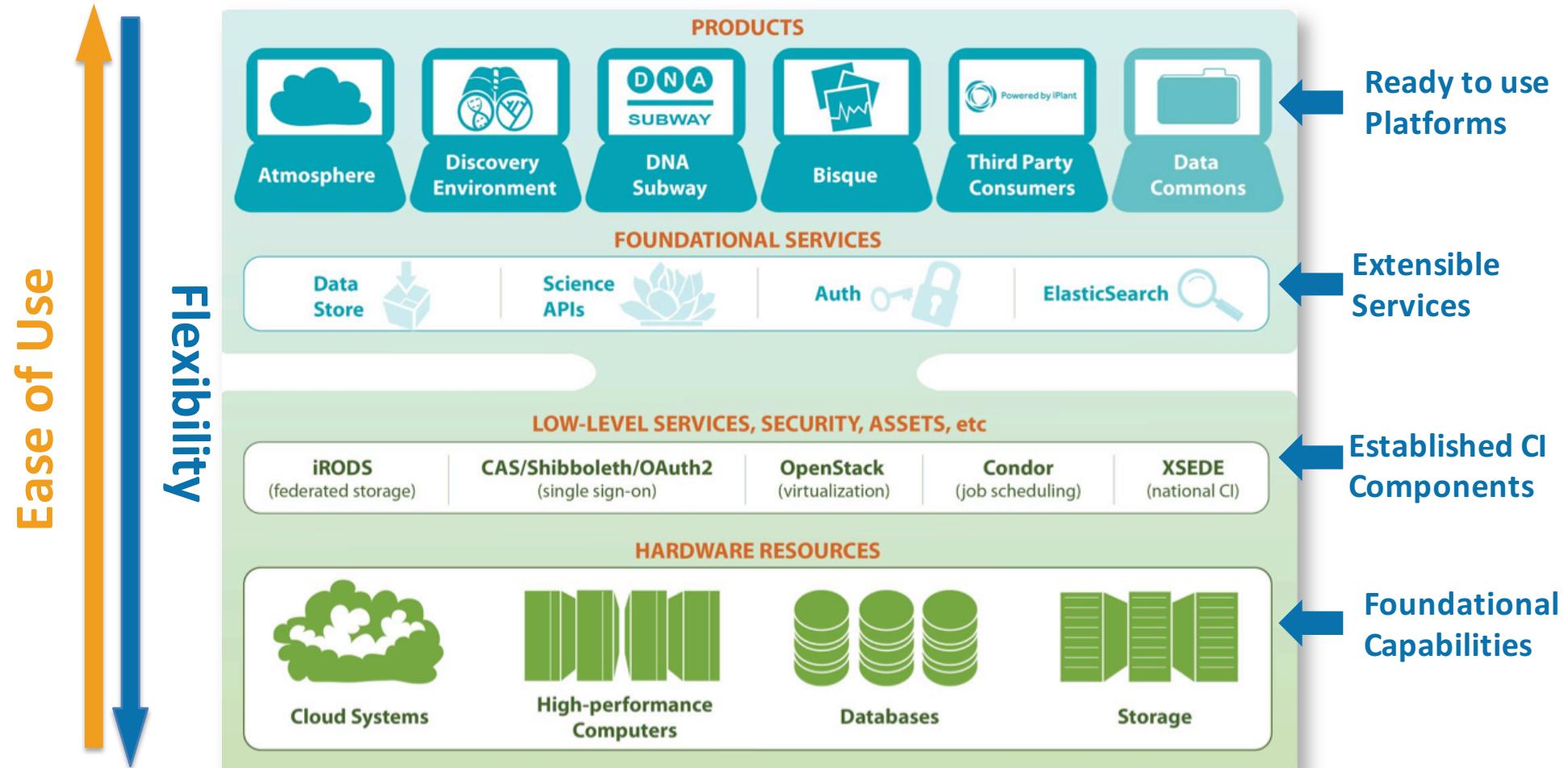


Ecological/Climate

CyVerse is built for Data



# CyVerse product stack



# How was CyVerse built?



Published online 3 September 2008 | *Nature* **455**, 16-21 (2008) | doi:10.1038/455016a

News Feature

# **Big data: Welcome to the petacentre**

**What** does it take to store bytes by the tens of thousands of trillio  
*Nature* 455, 30 (4 September 2008) | doi:10.1038/455030a; Published online 3 September 2008

## **which** Big data: Distilling meaning from data

Felice Frankel<sup>1</sup> & Rosalind Reid<sup>2</sup>

**Buried in va:  
need to craft  
Rosalind Rei** *Nature* 455, 28–29 (4 September 2008) | doi:10.1038/455028a; Published online 3 September 2008

## Big data: How do your data grow?

Clifford Lynch<sup>1</sup>

1. Clifford Lynch is the executive director of the Coalition for Networked Information, 21 Dupont Circle, Washington DC 20036, USA, and an adjunct professor at the School of Information, University of California, Berkeley, California, 94720-4600, USA.  
Email: cliff@cni.org

**Scientists need to ensure that their results will be managed for the long haul. Maintaining data takes big organization, says Clifford Lynch.**

[Purchase](#) | [455](#), 1 (4 September 2008) | doi:10.1038/nature01a; Published online 3 September 2008

Community cleverness required

**Researchers must learn to adapt institutions and practices in response to torrents of new data — and need to complement smart science with smart searching.**



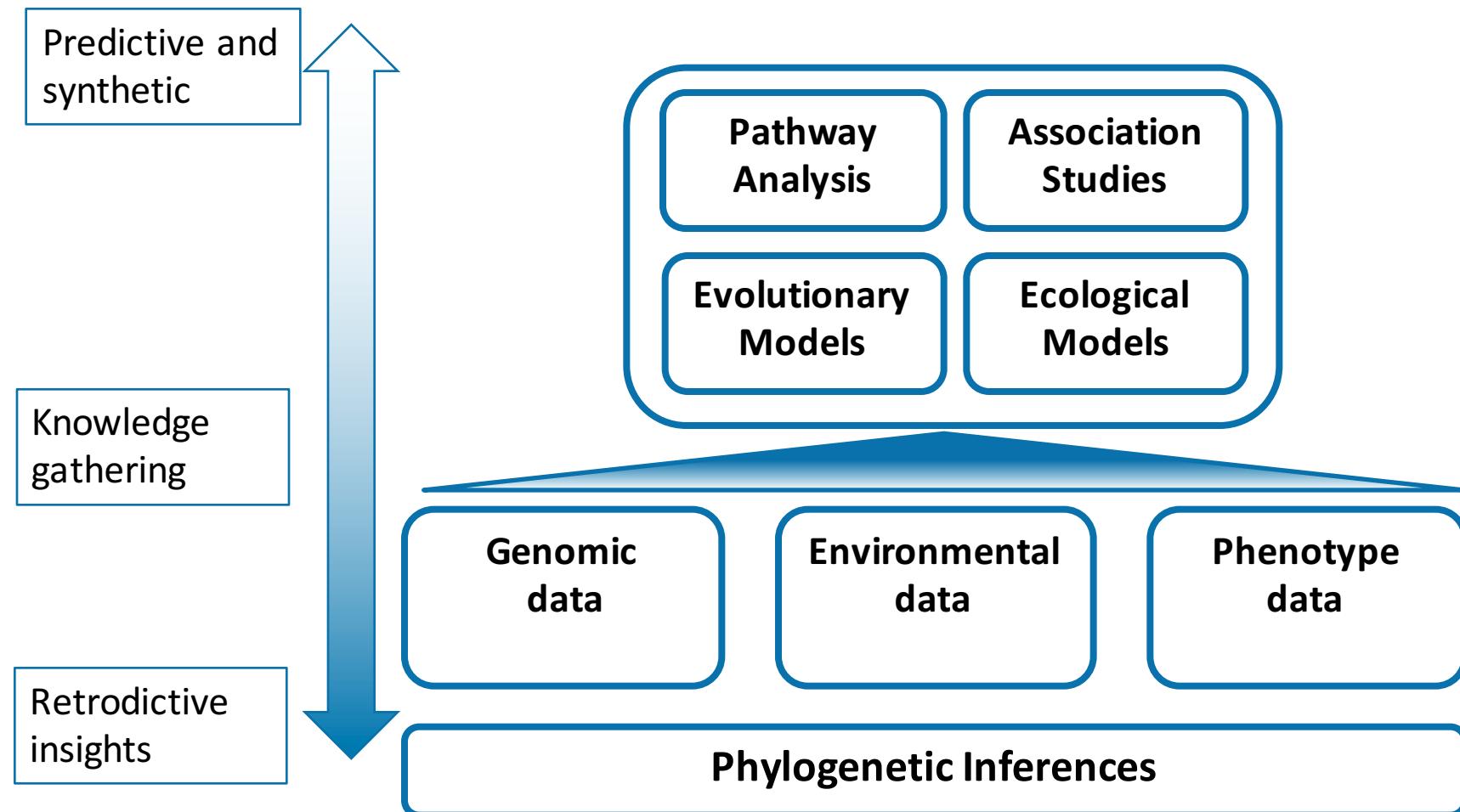
# CyVerse Community Priorities

## Genomic data and analysis:

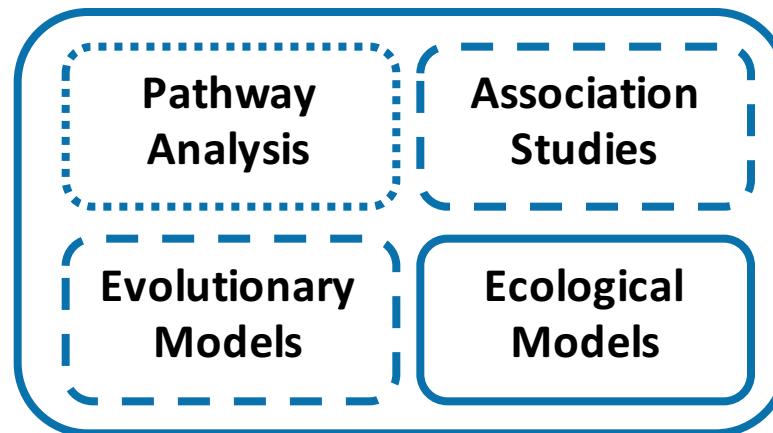
- Reference guided assembly
- De novo assembly
- RNA-Seq (expression; gene/isoform discovery)
- Variant calling
- Genome/Transcriptome annotation
- ChIP-Seq/Integration of epigenetic information
- Multiple sequencing platforms
- New and evolving technologies



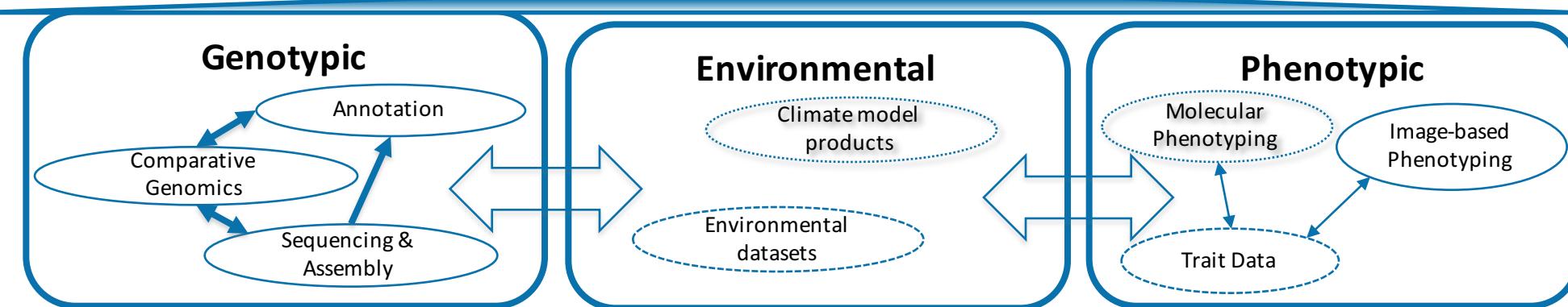
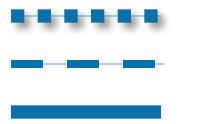
# CyVerse Community Priorities



# CyVerse Community Priorities



In planning  
In progress  
Foundation in place



Phylogenetic Tools for inference

# CyVerse Collaborators



DOE Systems Biology Knowledgebase



Arabidopsis Information Portal



 BioExtract Server  
data access, analysis, storage, and workflow creation



CyVerse collaborates to enable access to the solutions that work the best for you...



# CyVerse Institutions



Cold  
Spring  
Harbor  
Laboratory

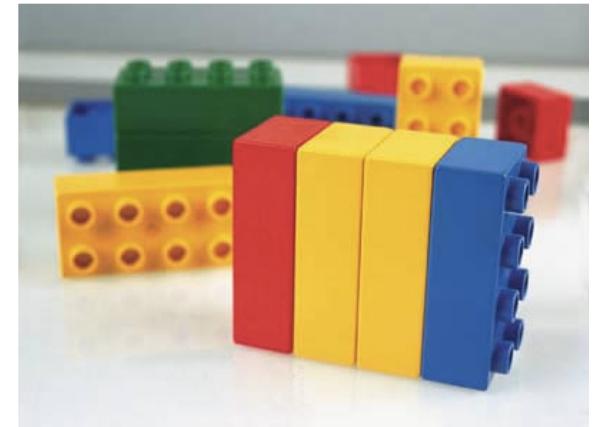


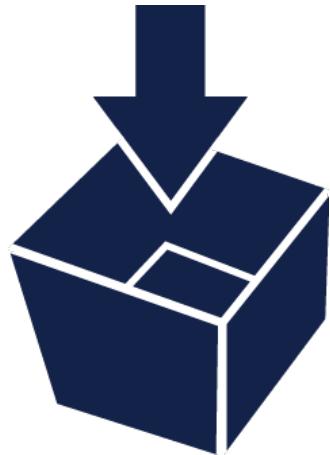
CyVerse is a collaborative virtual organization



# CyVerse Products

- We strive to be the **CI Lego blocks**
- Danish 'leg godt' - **'play well'**
- Also translates as '**I put together**' in Latin
- If a solution is not available you can craft your own using CyVerse CI components





# Data Store

The resources you need to share and manage data with your lab, colleagues and community

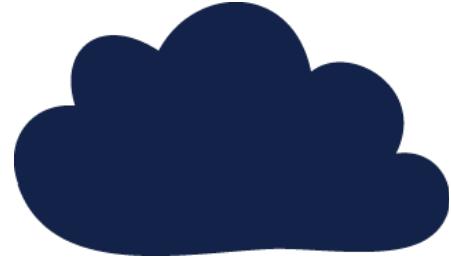
- ✓ Initial 100 GB allocation – TB allocations available
- ✓ Automatic data backup
- ✓ Easy upload /download and sharing



# Discovery Environment

Hundreds of bioinformatics Apps in an easy-to-use interface

- ✓ A platform that can run almost any bioinformatics application
- ✓ Seamlessly integrated with data and high performance computing
- ✓ User extensible – add your own applications



# Atmosphere

Cloud computing for the life sciences

- ✓ Simple: One-click access to more than 200 virtual machine images
- ✓ Flexible: Fully customize your software setup
- ✓ Powerful: Integrated with iPlant computing and data resources





# Science APIs

Fully customize *iPlant* resources

- ✓ Science-as-a-service platform
- ✓ Define your own compute, and storage resources (local and *iPlant*)
- ✓ Build your own app store of scientific codes and workflows





# DNA Subway

Educational workflows for Genomes, DNA Barcoding, RNA-Seq

- ✓ Commonly used bioinformatics tools in streamlined workflows
- ✓ Teach important concepts in biology and bioinformatics
- ✓ Inquiry-based experiments for novel discovery and publication of data





# Bisque

Image analysis, management, and metadata

- ✓ Secure image storage, analysis, and data management
- ✓ Integrate existing applications or create new ones
- ✓ Custom visualization and image handling routines and APIs





# CYVERSE™

Transforming Science Through Data-driven Discovery

## CyVerse Executive Team



THE UNIVERSITY  
OF ARIZONA.

**Parker Antin**  
**Nirav Merchant**  
**Eric Lyons**



TEXAS ADVANCED  
COMPUTING CENTER

**Matt Vaughn**



Cold  
Spring  
Harbor  
Laboratory

**Doreen Ware**  
**Dave Micklos**

CyVerse is supported by the National Science Foundation under Grant No. DBI-0735191 and DBI-1265383.

