iPlant:
Supporting the Lifecycle of Data

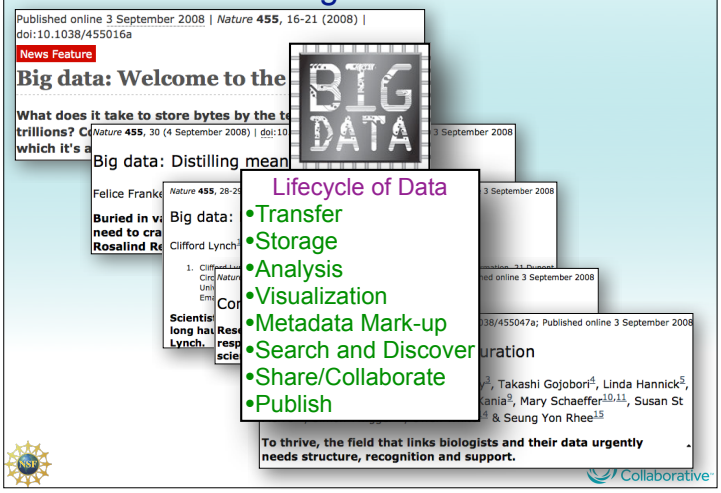
Eric Lyons

One Big Problem...

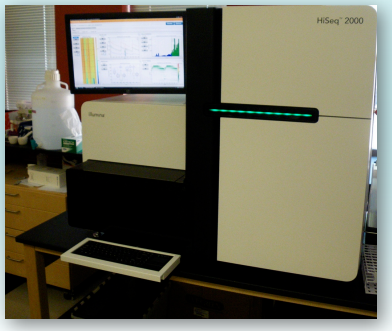
2008

2011

One Big Problem...

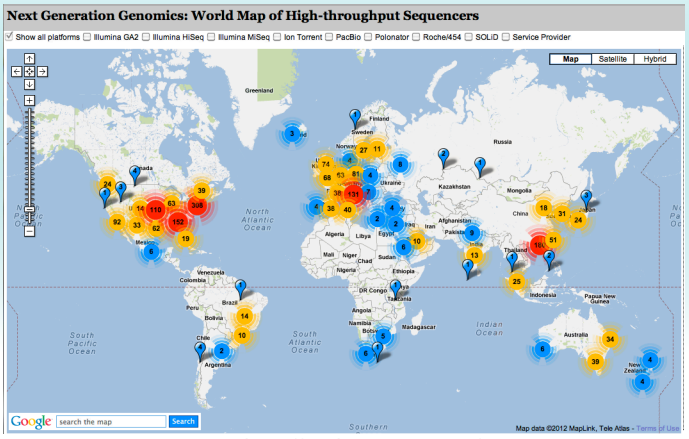


High-throughput Data Acquisition



In 11 Days
- Generates 4TB of raw data
- 600,000,000,000 bases of DNA sequence (200 human genomes)

and it is a global phenomenon



http://omicsmaps.com/

High-throughput Phenotyping
(Watching Grass Grow)

- $70K for ~30 camera sets
- ~200 movies of plants undergoing a dynamic growth process
- "Only" 4GB a day

Big Data in Ecology

Global Multidimensional Data

iPlant Collaborative

Lifecycle of Data

Biologist

- Transfer
- Storage
- Analysis
- Visualization
- Metadata Mark-up
- Search and Discover
- Share/Collaborate
- Publish

iPlant Collaborative

iPlant Data Store

Biologist

- Transfer
- Storage
- Analysis
- Visualization
- Metadata Mark-up
- Search and Discover
- Share/Collaborate
- Publish

iPlant Collaborative

The journey of data to iDS and challenges along the way !

Hard Drive → Network card → Building network → Campus network → Internet

Internet → UA/TACC network → iDS → Network card → Hard Drive

http://en.wikipedia.org/wiki/List_of_device_bandwidths

Check: USB, HDD, Network capabilities

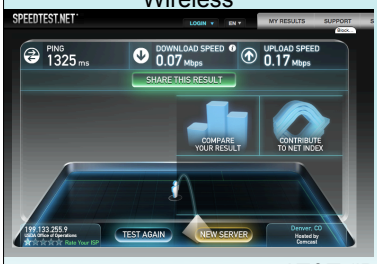
iPlant Collaborative

The Flow of Data

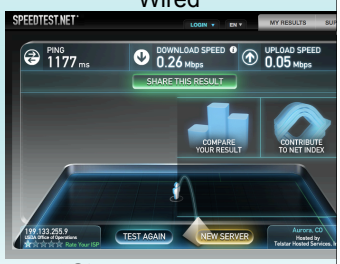
iPlant Collaborative

How fast is our network?

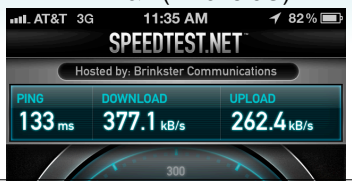
Wireless



Wired

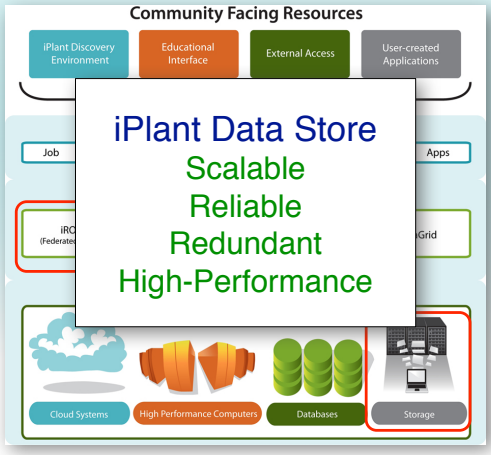


AT&T (iPhone 3G)



iPlant Collaborative

iPlant Layered Services and Access

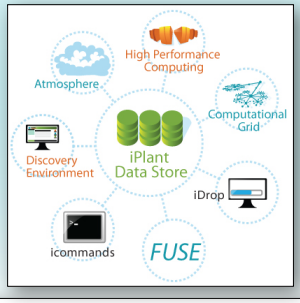


iPlant Data Store
Scalable
Reliable
Redundant
High-Performance

End Users

Computational Users

iPlant Data Store
Free Your Data

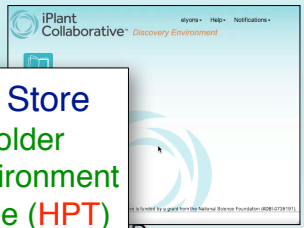


Different Users,
Different Access Needs:
One Data Store

iPlant Data Store
Free Your Data

WebDAV

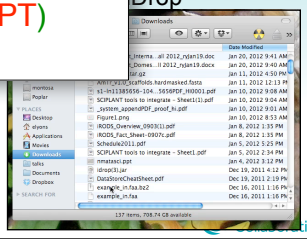
DE



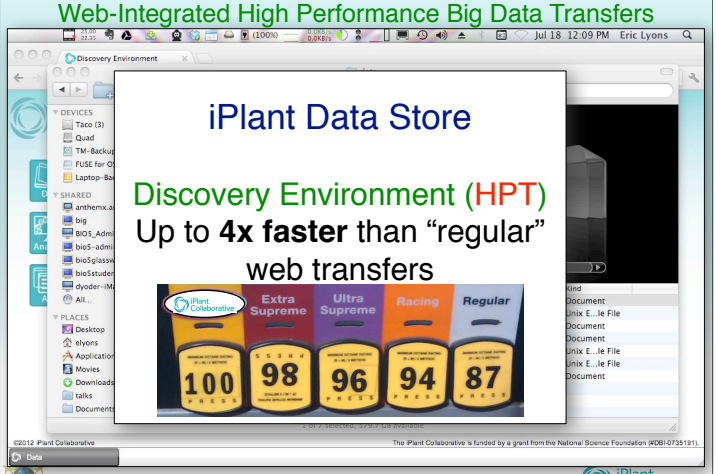
iPlant Data Store
Desktop Folder
Discovery Environment
Command Line (HPT)
i-Drop (HPT)
API

i-comman

Drop



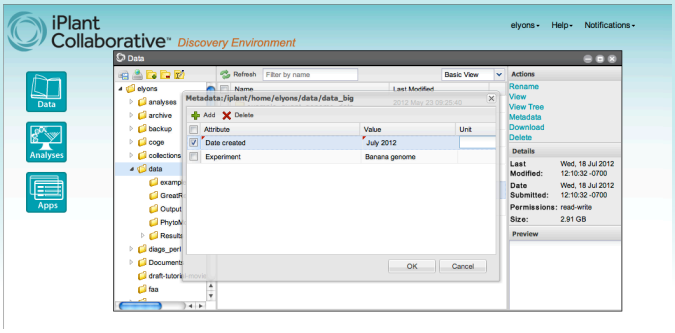
iPlant Data Store
Web-Integrated High Performance Big Data Transfers



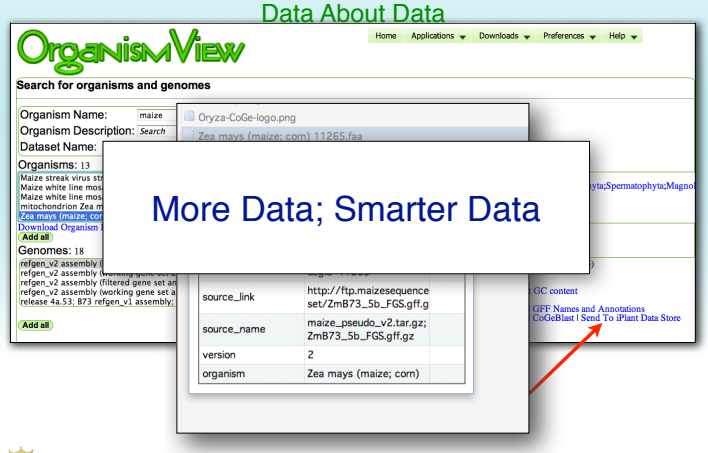
iPlant Data Store

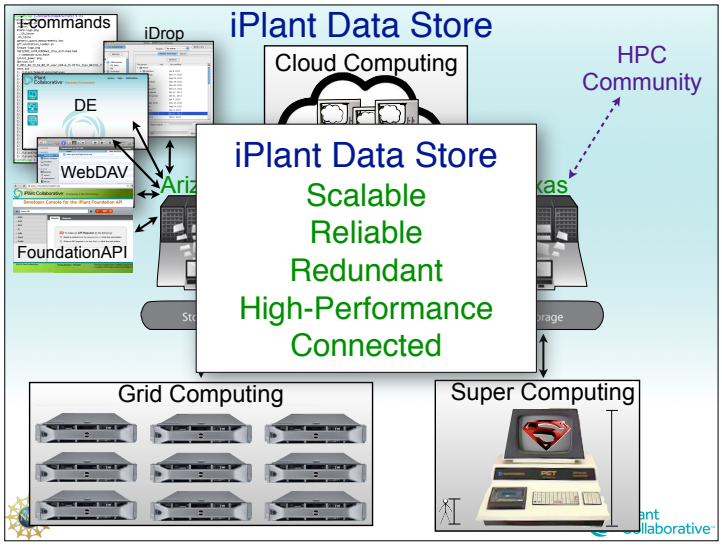
Discovery Environment (HPT)
Up to **4x faster** than "regular" web transfers

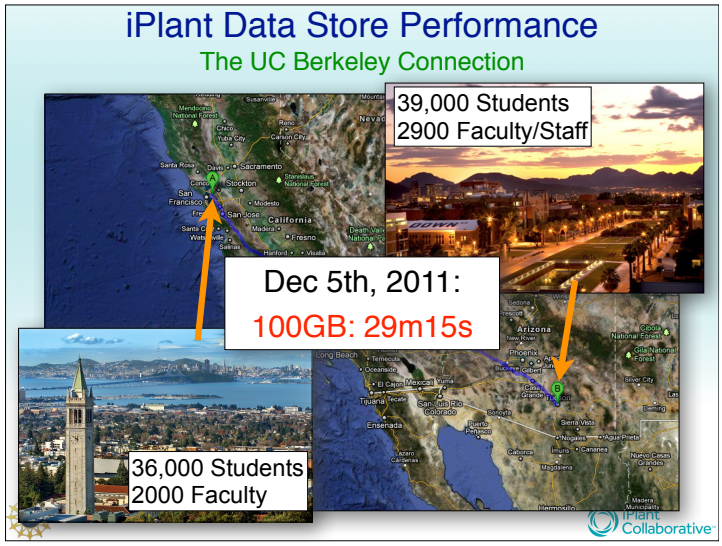
iPlant Data Store: Metadata
Data About Data



iPlant Data Store: Metadata
Data About Data







iPlant Data Store Performance

The UC Berkeley Connection

100GB: 29m15s

1 GB / 17.5 seconds

Source	Destination	Copy Method	Time (seconds)
CD	My Computer	cp	320
Berkeley Server	My Computer	scp	150
External Drive	My Computer	cp	36
USB2.0 Flash	My Computer	cp	30
iDS	**My Computer**	**iget**	**18**
My Computer	My Computer	cp	15

My Computer (UA): 7.2K Internal Hard Drive
External Drive: USB 2.0: 5.4k Hard Drive
Flash Drive: USB2.0 Patriot XT

https://pods.iplantcollaborative.org/wiki/display/start/How+fast+is+the+iPlant+Data+Store

iPlant Data Store

Connecting people with data and computation:

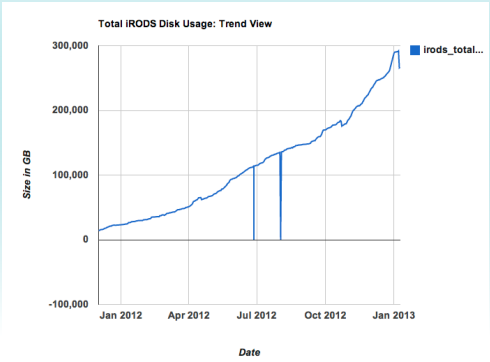
Lifecycle of Data

Powered By iPlant

Cyberinfrastructure for Life Sciences

Scalable
Capable
Extensible

Growing about 1TB/day:



Where to Get Information

Data Store Quick Start:
http://www.iplantcollaborative.org/Zki

Data Store Manual:
http://www.iplantcollaborative.org/Zko

iPlant Forums:
http://forums.iplantcollaborative.org

Exercise Your Data Store

For each (WebDAV, DE, iDROP, i-command):

{

1. Upload your data
2. Rename your data
3. Delete your data

}

iPlant Collaborative