

Galaxy

Introduction to the Galaxy platform

www.galaxyproject.org

Galaxy: accessible analysis system

The screenshot displays the Galaxy web interface in a browser window. The address bar shows the URL <http://main.g2.bx.psu.edu/>. The top navigation bar includes links for **Analyze Data**, **Workflow**, **Data Libraries**, **Admin**, **Help**, and **User**.

Tools Panel (Left): A list of available tools categorized into sections like **Get Data**, **Send Data**, **ENCODE Tools**, **Lift-Over**, **Text Manipulation**, **Convert Formats**, **FASTA manipulation**, **Filter and Sort**, **Join, Subtract and Group**, **Extract Features**, **Fetch Sequences**, **Fetch Alignments**, **Get Genomic Scores**, **Operate on Genomic Intervals**, **Statistics**, **Graph/Display Data**, **Regional Variation**, **Multiple regression**, **Multivariate Analysis**, **Evolution**, **Metagenomic analyses**, **EMBOSS**, **NGS TOOLBOX BETA**, **NGS: QC and manipulation**, **NGS: Mapping**, **NGS: SAM Tools**, **NGS: Peak Calling**, **RGENETICS**, **SNP/WGA: Data; Filters**, and **SNP/WGA: QC; LD; Plots**.

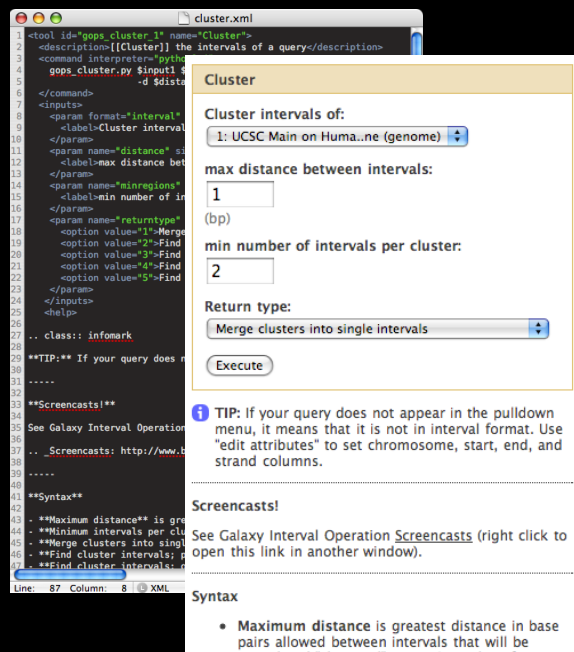
Main Content Area: A central panel titled "Here is what's happening..." features a large box with the text "Mapping Pipeline for Illumina, 454, and SOLiD" and a prominent "USE IT NOW!" button. Below this, a section titled "Live Quickies (more after May 17 ...)" displays three cards: "Basic fastQ manipulation: Galactic quickie # 13", "Advanced fastQ manipulation: Galactic quickie # 14", and "454 Mapping: Single End: Galactic quickie # 15". At the bottom of the main area, text states: "The Galaxy team is a part of BX at Penn State. This project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, and The Institute for CyberScience at Penn State. Galaxy build: \$Rev 3885:1ab9d6b0ddfc\$".

History Panel (Right): A list of previous analyses, each with a name, an eye icon, a refresh icon, and a delete icon. The list includes: "imported: metagenomic analysis", "16: Draw phylogeny on data 14", "15: Summarize taxonomy on data 13", "14: Find lowest diagnostic rank on data 13", "13: Fetch taxonomic representation on data 12", "12: Filter on data 11", "11: Join two Queries on data 9 and data 10", "10: Concatenate queries on data 8 and data 7", "9: Compute sequence length on data 6", "8: Megablast on data 6", "7: Megablast on data 6", "6: Tabular-to-FASTA on data 5", "5: Add column on data 4", and "4: FASTA-to-Tabular on data 4".

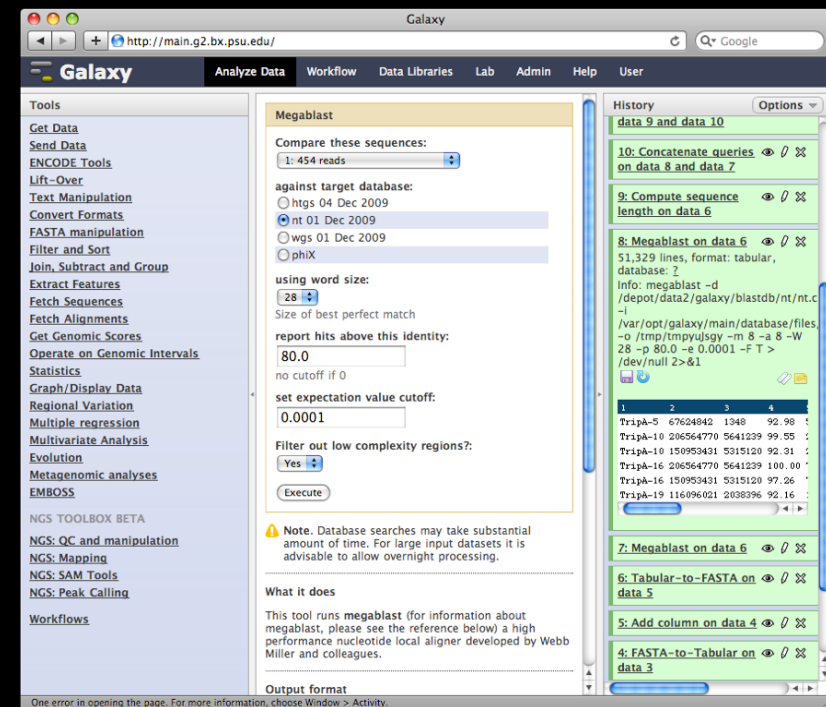
A free (for everyone) web service integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage

Open source software that makes integrating your own tools and data and customizing for your own site simple

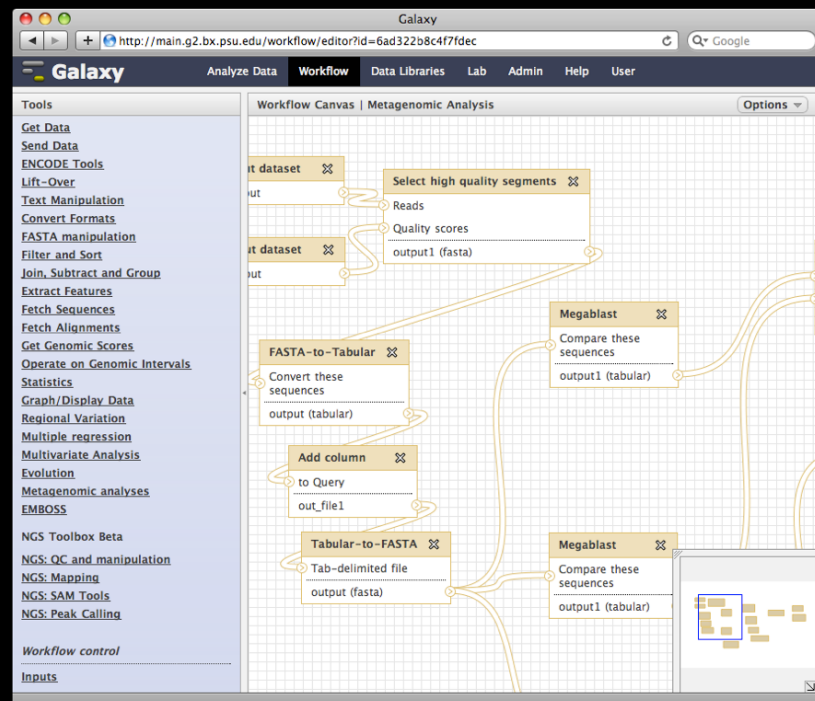
An open extensible platform for sharing tools, datatypes, workflows, ...



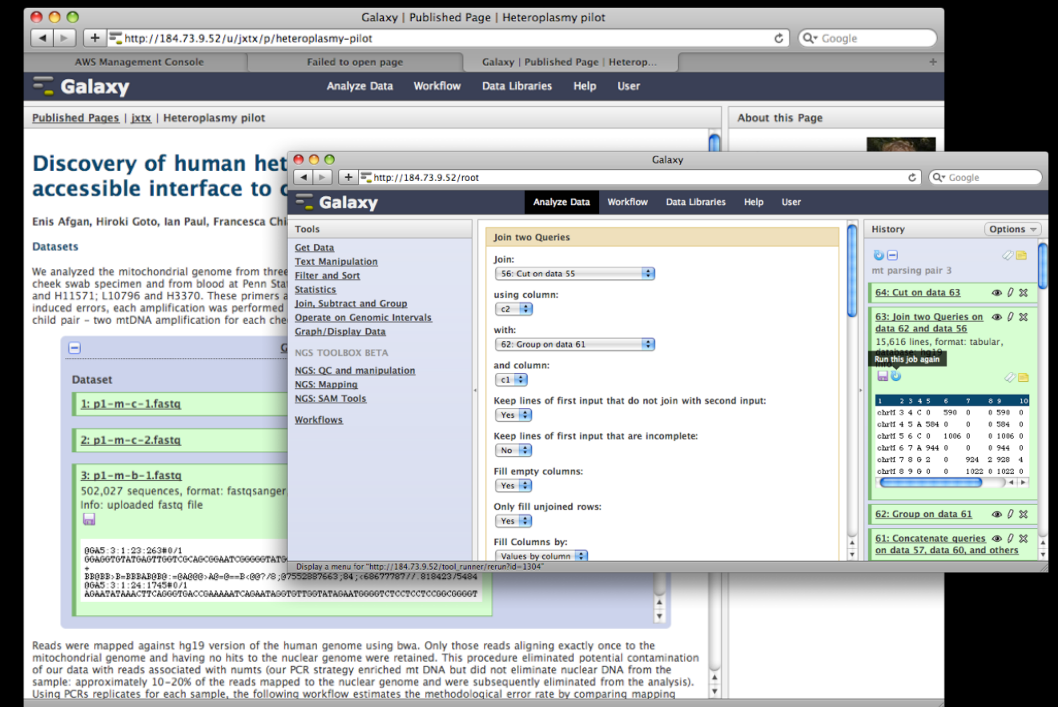
Describe analysis tool
behavior abstractly



Analysis environment automatically
and transparently tracks details



Workflow system for complex analysis,
constructed explicitly or automatically



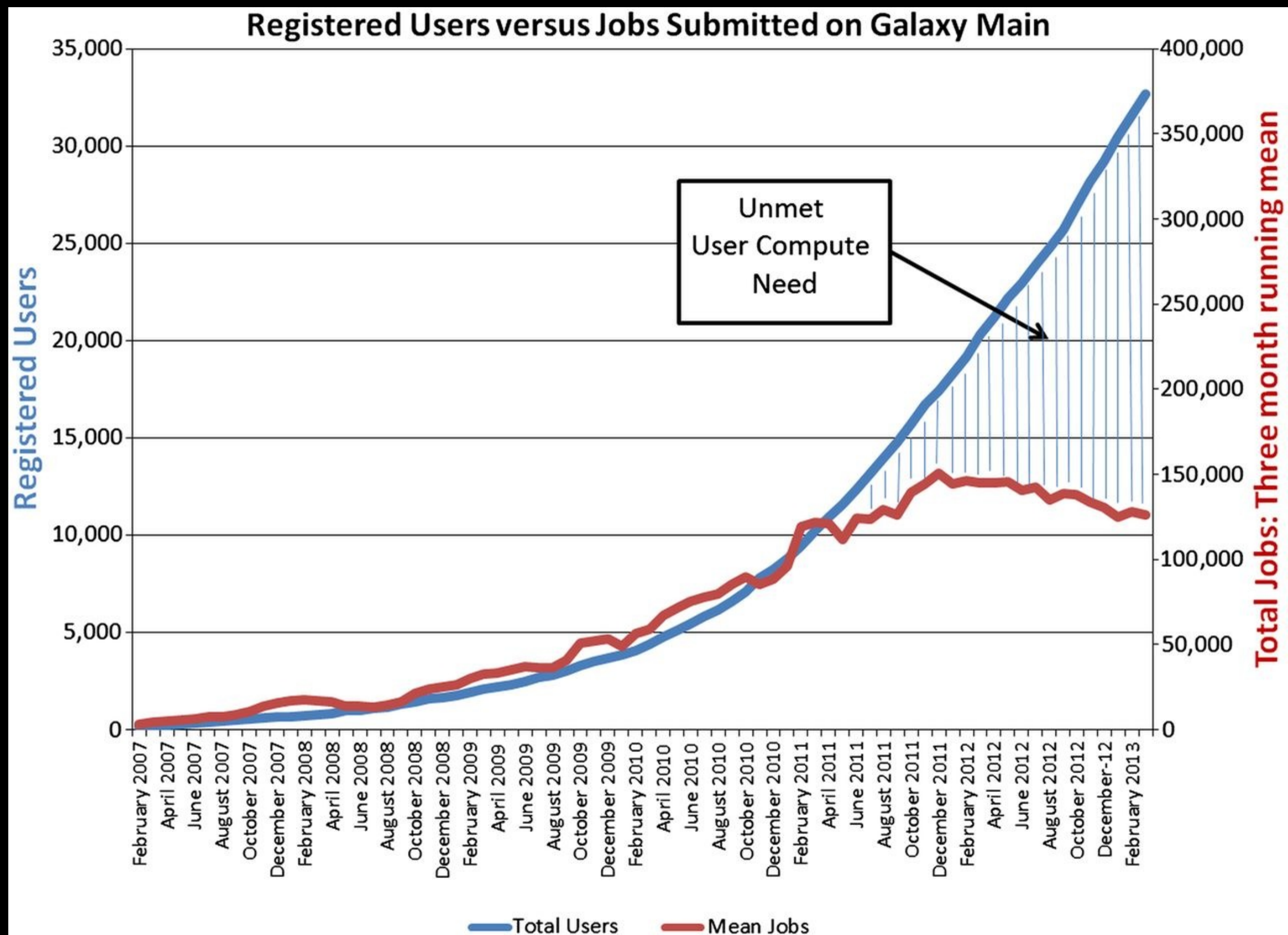
Pervasive sharing, and publication
of documents with integrated analysis

Ways to use Galaxy...

Galaxy is available as a free (for everyone) web server integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage:

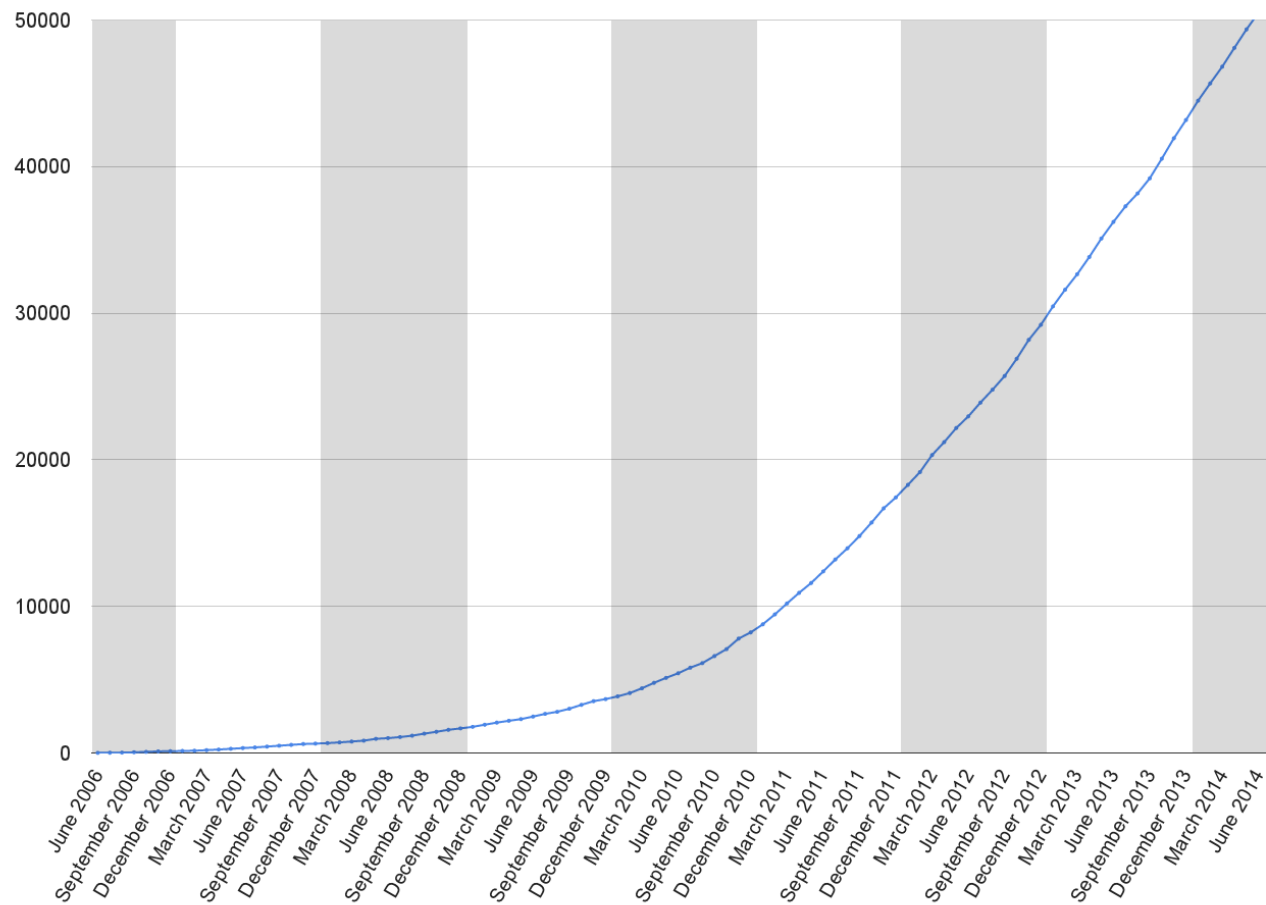
<http://usegalaxy.org>

However, *a centralized solution cannot support the different analysis needs of the entire world!*



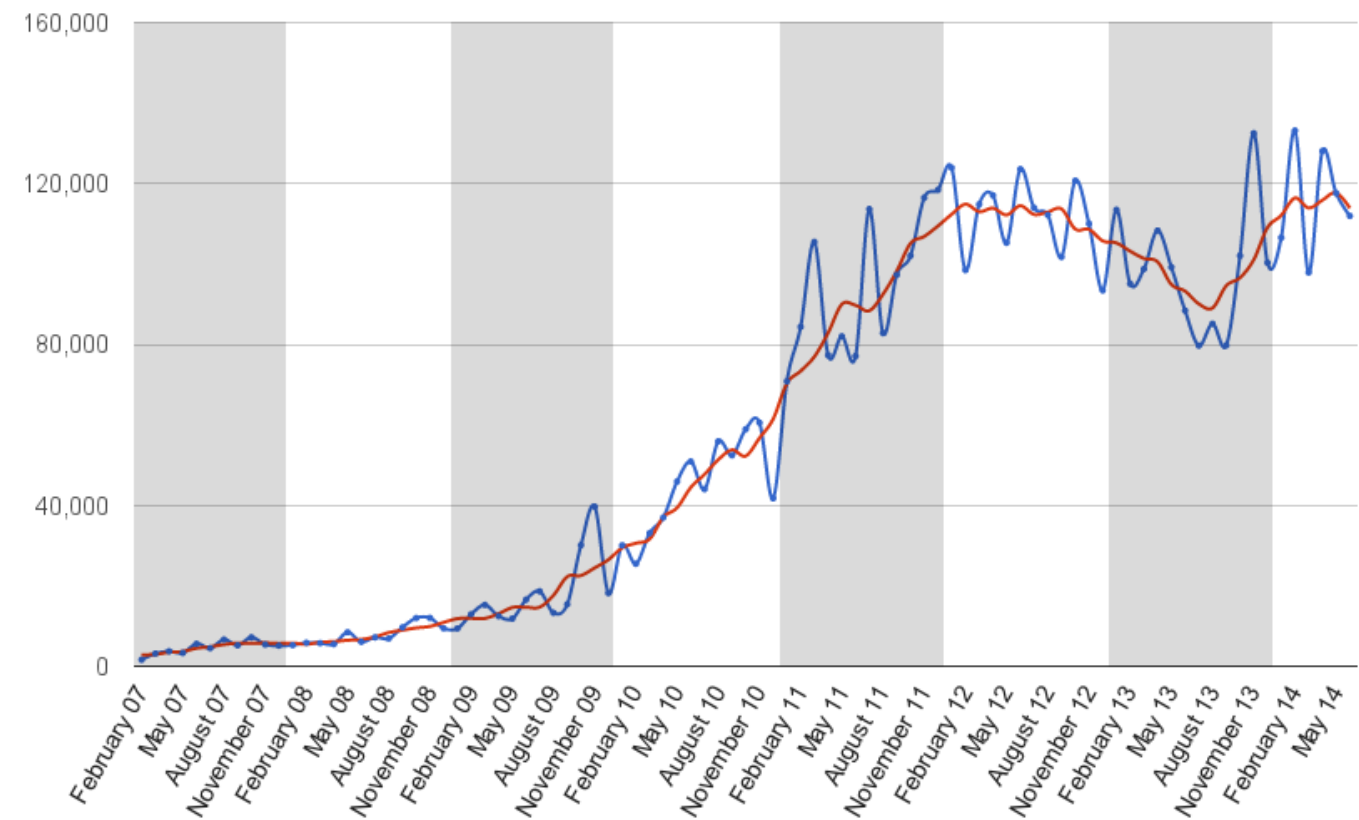
Leveraging the national cyberinfrastructure for biomedical research
LeDuc, et al. *J Am Med Inform Assoc* doi:10.1136/amiajnl-2013-002059

Registered Users on Galaxy Main



Jobs on usegalaxy.org

1 Month 6 Month Avg



And those trends continue...

bit.ly/gxyStats

Alternatives to the public service

- 1) Local installation
- 2) Cloud Computing

Local Galaxy Deployment

Galaxy is designed for local installation and customization... just download and run

Pluggable interfaces to compute resources, easily connect to one or more existing clusters

Ideally, allow users to take advantage of whatever computational resources they already have access to.

More than 60 known public Galaxy servers

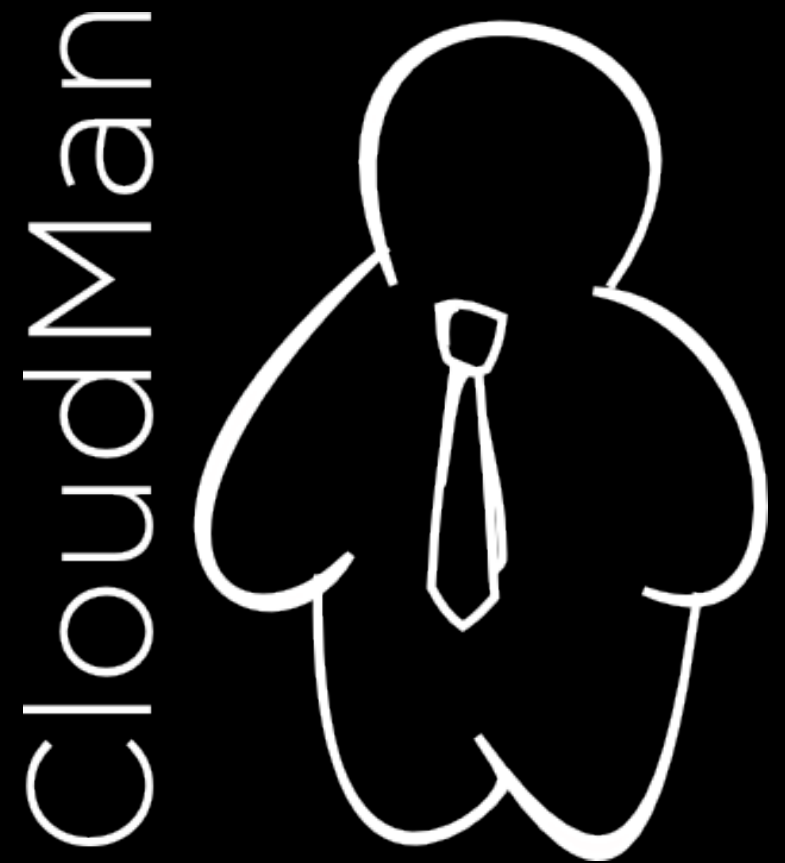
Ballaxy for structure based computational biology,
Cistrome for regulatory sequence analysis,
Genomic Hyperbrowser: statistical integration of genomic data,
GigaGalaxy: integrating workflows published in GigaScience,
Pathogen Portal: comparative analysis of host response to pathogens,
...

Dozens of large scale private Galaxy instances



Welcome to **Galaxy** on the Cloud(s)

CloudMan: a general
purpose deployment
manager for **ANY** cloud



Summary

Reproducibility is crucial to computational research,
but currently poorly implemented

Galaxy is one platform that seeks to make
computational analysis more *accessible, transparent,*
and *reproducible*

It can be used as a free web service ([http://
usegalaxy.org](http://usegalaxy.org)), installed locally, or deployed in a
cloud environment