

# State of the Galaxy

(Wie läuft's?)

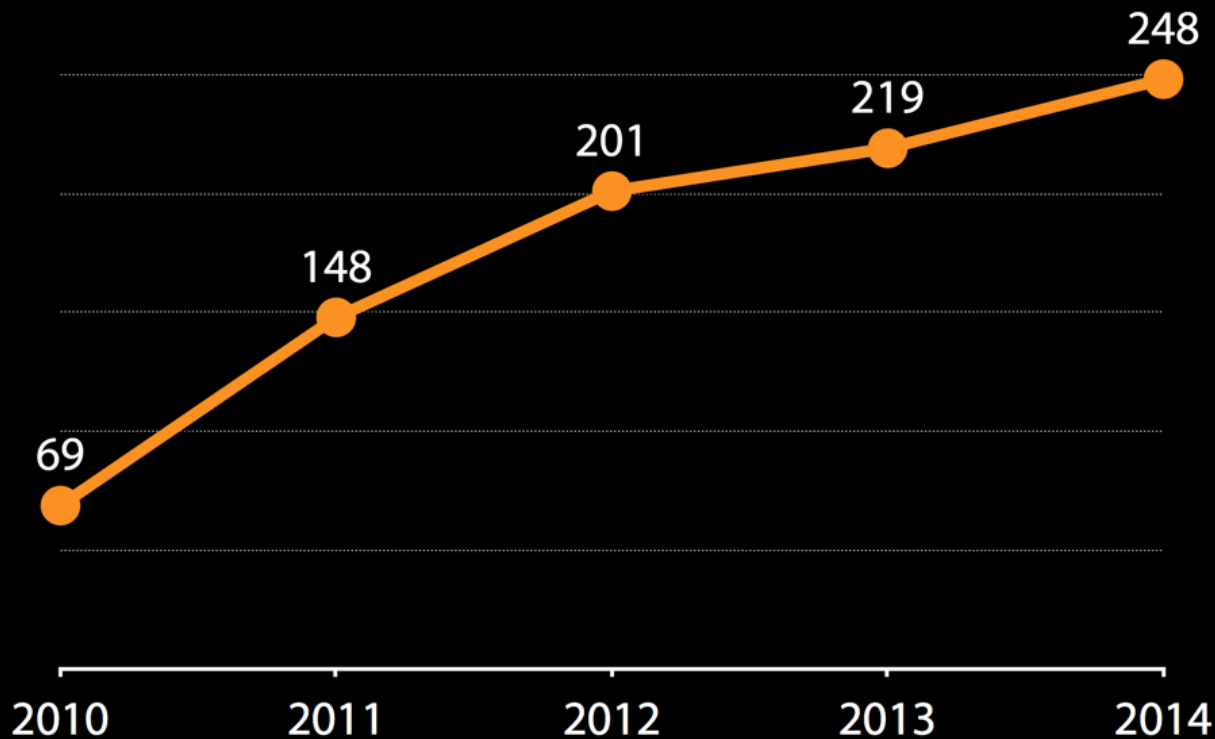
Dannon Baker, Johns Hopkins University,  
Galaxy Team

1. **Community** — GCC, public instances, code and tool contributions, publications
2. **Galaxy Main** — TACC move(more detail), Galaxy Main usage
3. **Recent Highlights** — pulsar, biostar, viz, dataset collections, toolshed, data managers, UI evolution
4. **Ongoing Efforts** — priority areas and things we're excited about

# State of the Community

GCC, public instances, code and tool  
contributions, papers

# Galaxy Community Conference



# Galaxy Community Conference



# Galaxy Hackathon: Making Connections, Providing Footholds

- Developer Community Building
- Bandwidth is important
- Great opportunity for 'fun' projects



# Hackathon Template Slide (copy me)

- Bullets
- Orange

Feel free to include screenshots/pictures.

Please do include the names of people who hacked with you on the project. Pictures and logos too, if you have them!

(contributor 1\*, contributor 2, contributor 3, docker)

[wiki.galaxyproject.org/PublicGalaxyServers](http://wiki.galaxyproject.org/PublicGalaxyServers)

14 General Purpose Servers

21 Domain Servers

30 Tool Publishing Servers

4 Public Toolsheds





# Code Contributions

## In a Nutshell, Galaxy Bioinformatics Platform...

- ... has had 14,841 commits made by 106 contributors representing 429,567 lines of code
- ... is mostly written in JavaScript with an average number of source code comments
- ... has a well established, mature codebase maintained by a very large development team with increasing Y-O-Y commits
- ... took an estimated 116 years of effort (COCOMO model) starting with its first commit in November, 2006 ending with its most recent commit 4 days ago

Send a pull request!

## 12 Month Summary

Sep 27 2013 — Sep 27 2014

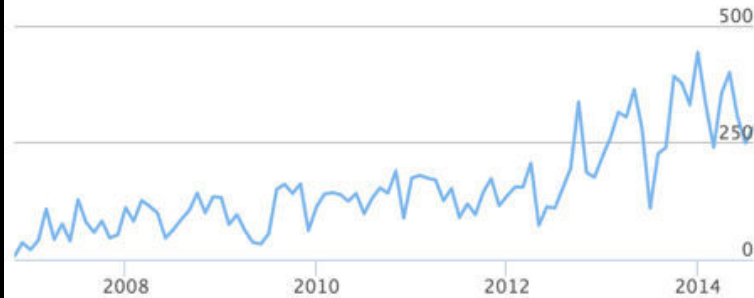
### 4029 Commits

Up +989 (32%) from previous 12 months

### 47 Contributors

Up +1 (2%) from previous 12 months

## Commits per Month



# The Galaxy Toolshed

897 repositories

222 unique owners

176 Tool dependency package installation recipes 2,330  
valid tools

3,420 valid versions of tools

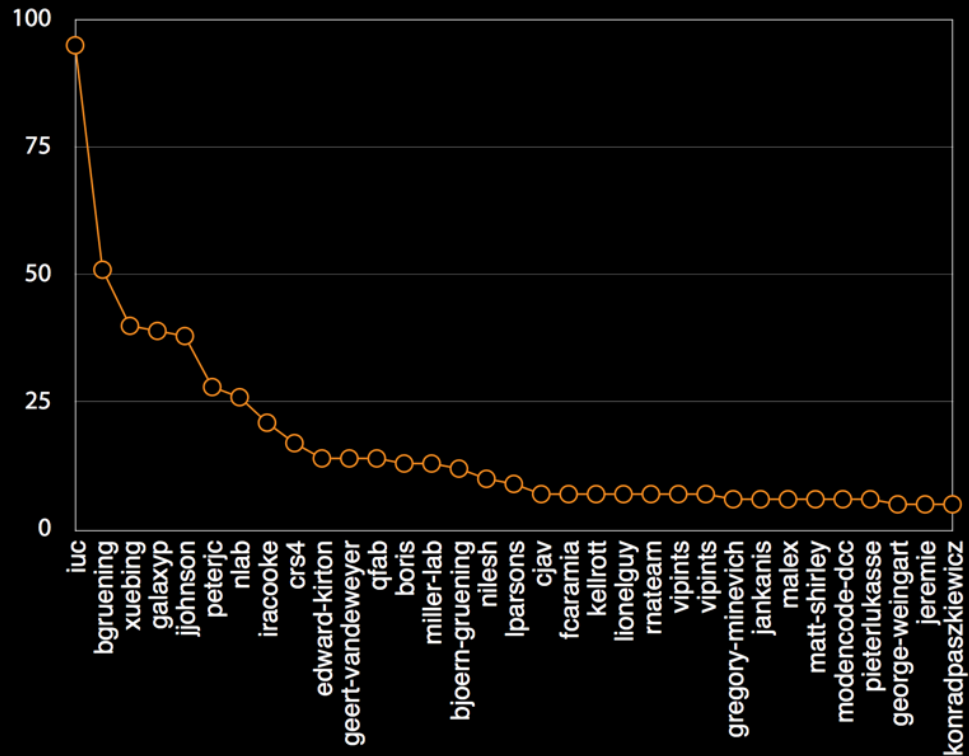
54 exported Galaxy workflows

455 custom datatypes

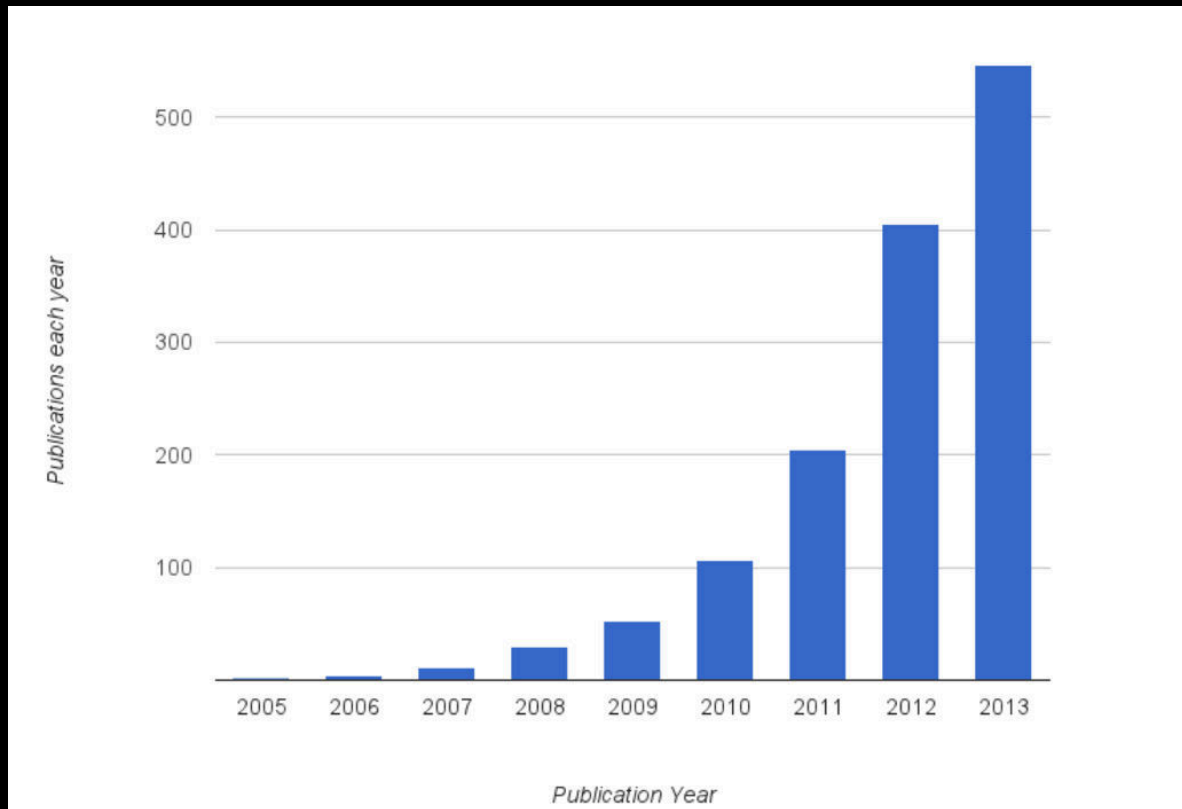
62,021 total repository installations

(as of July 2014)

# Toolshed Contributions

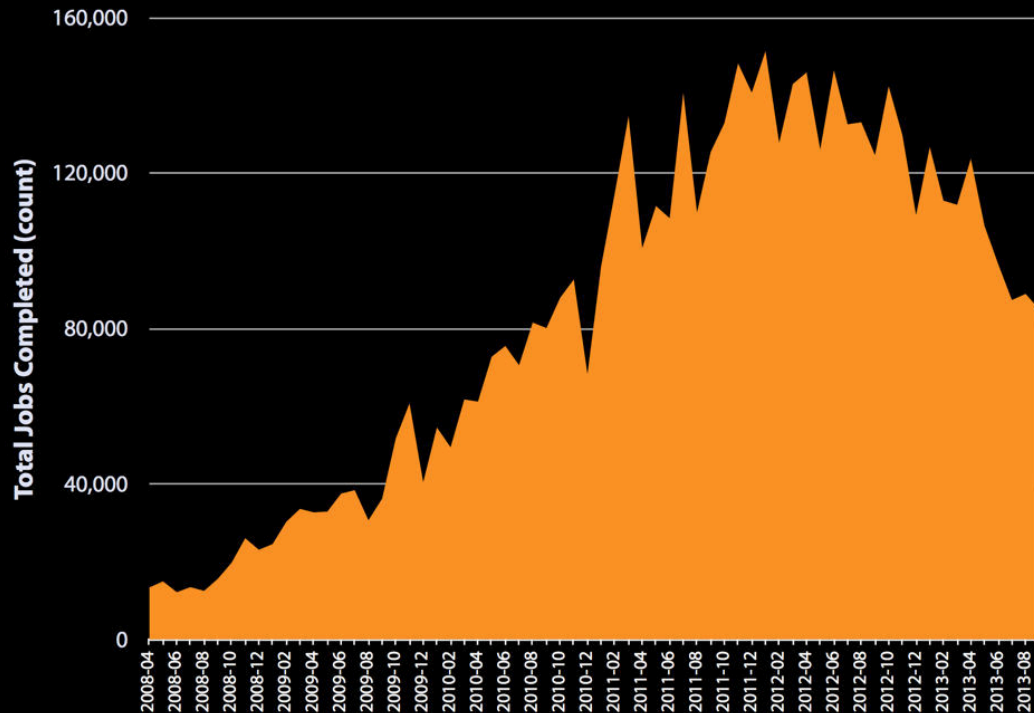


# Publications (this just got even easier)

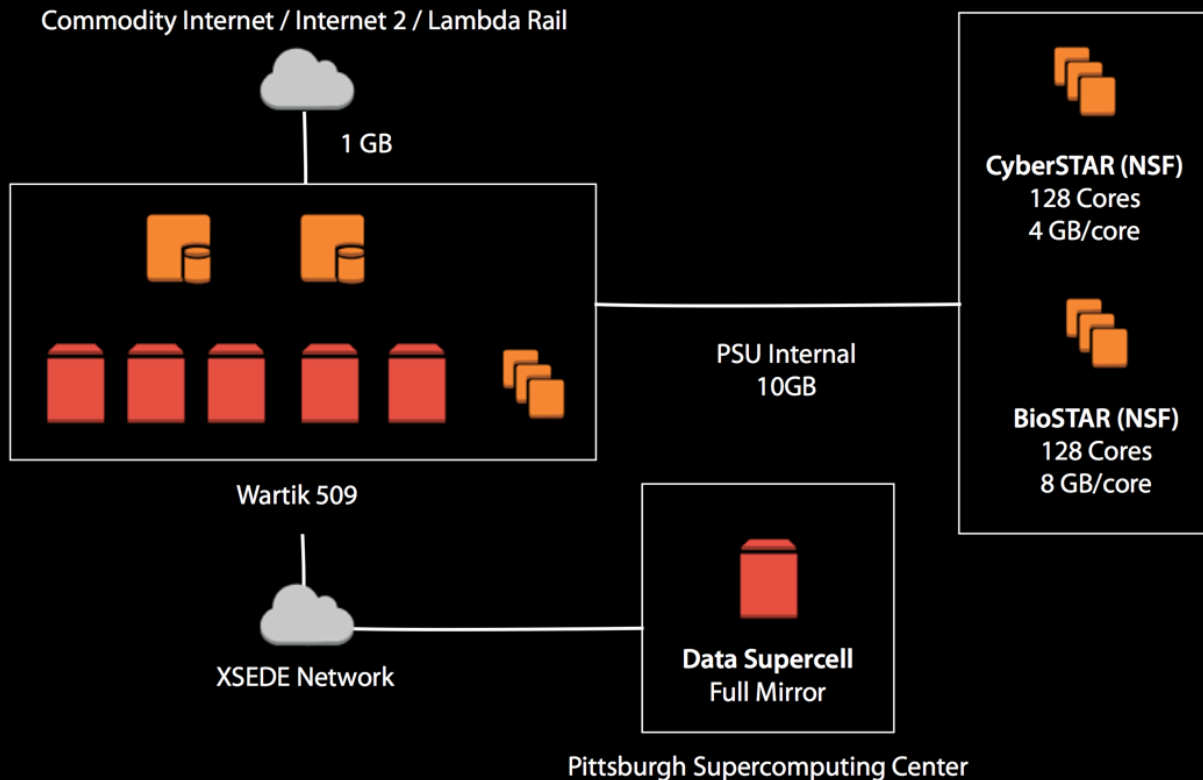


# The State of Galaxy Main

# Galaxy Main (usegalaxy.org)



# Architecture during 'job numbers crash'

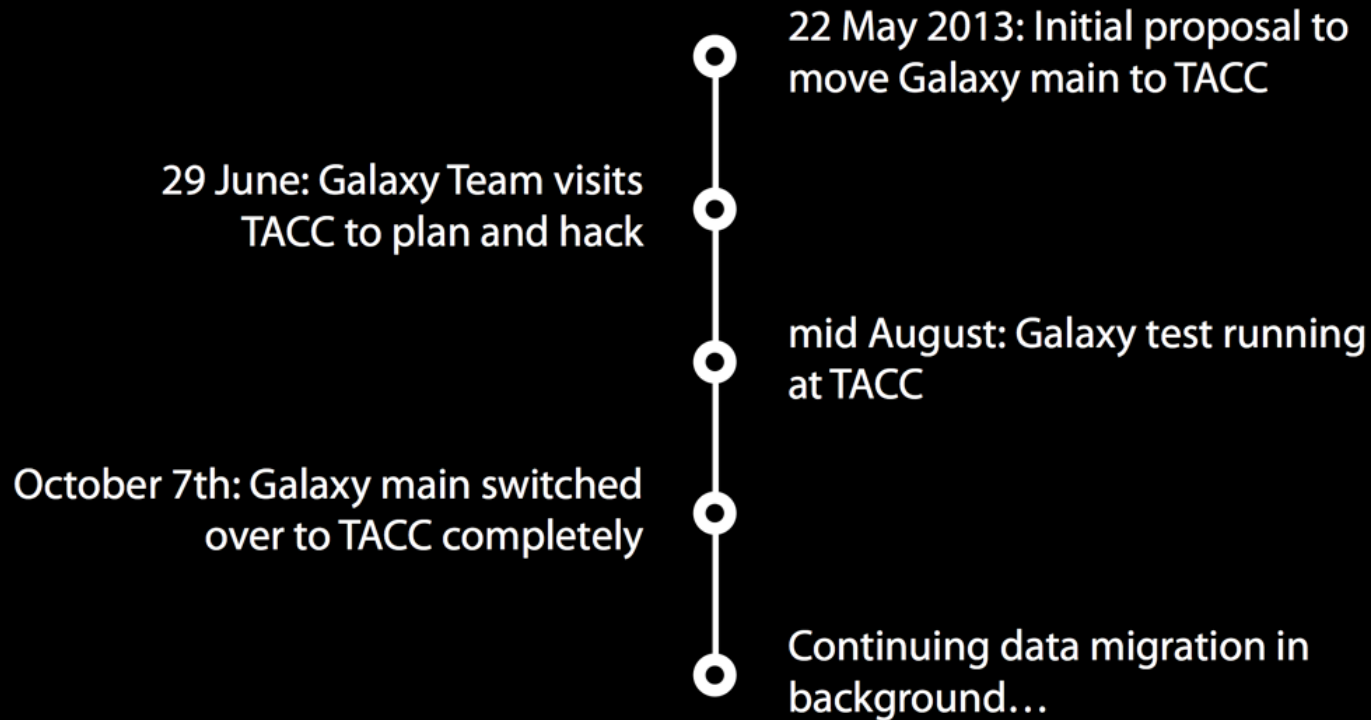


# Galaxy Main ([usegalaxy.org](https://usegalaxy.org))

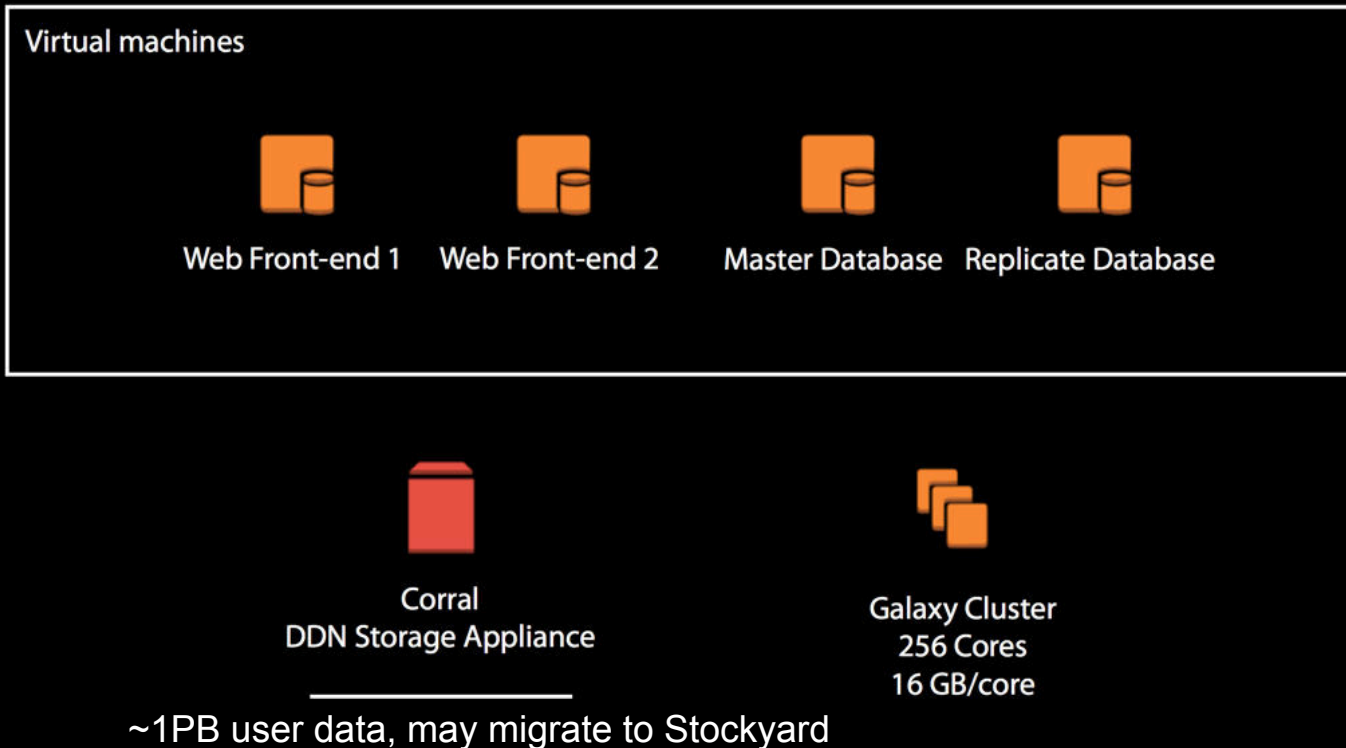
Quotas: (250/5 GB)

Concurrent Job Limiting: (6/1)

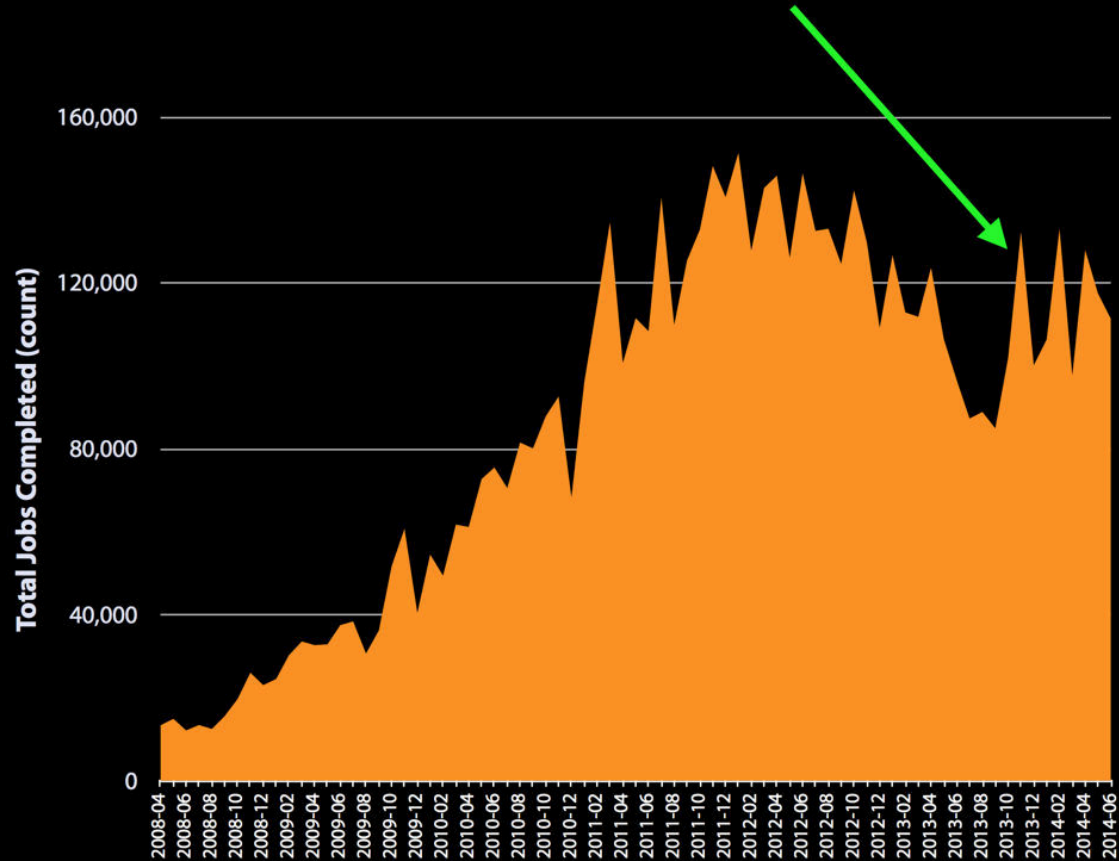




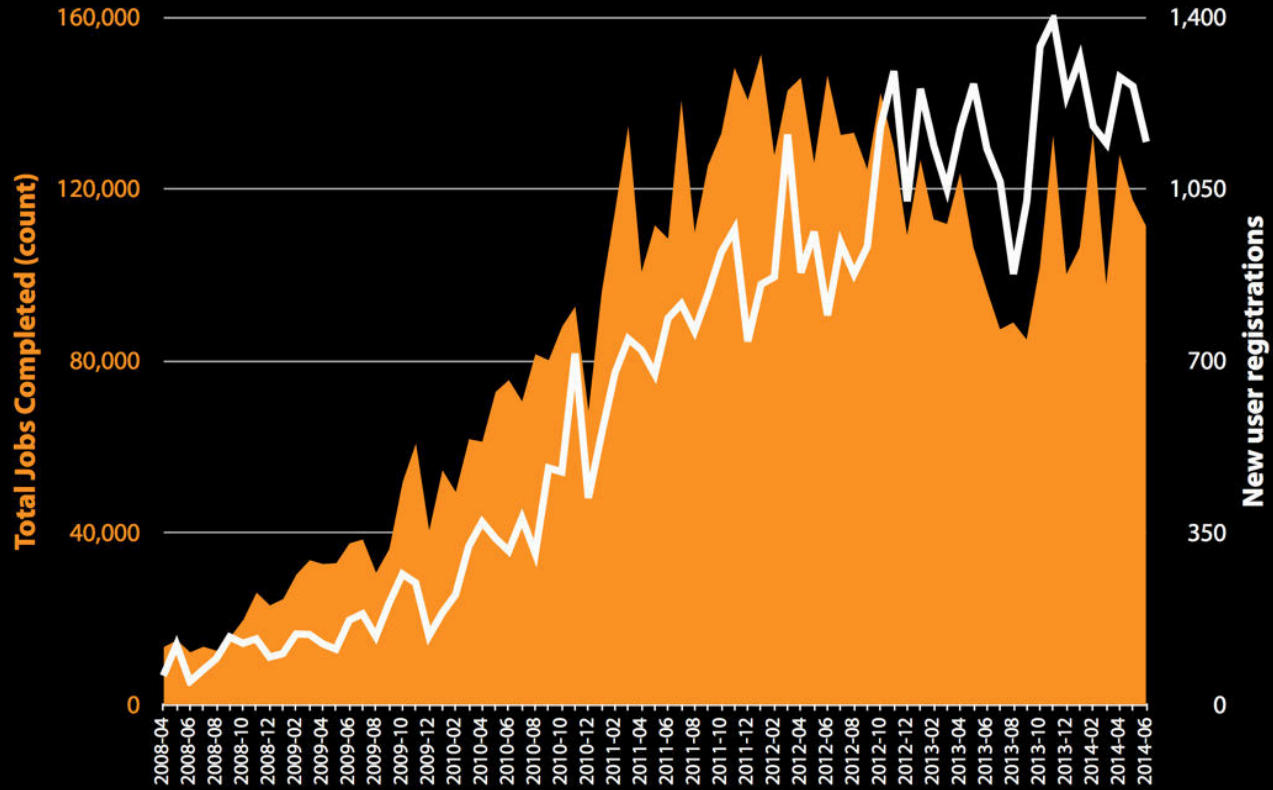
# Galaxy Main at TACC



# Job throughput post-move



# usegalaxy.org userbase



# Recent Highlights

# Ansible for automated deployments

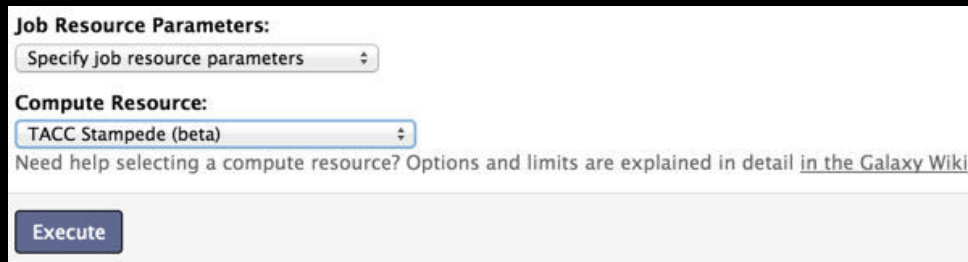
- Started as our private playbook for automating changes to main, galaxyproject.org, etc.
- [github.com/galaxyproject/ansible-galaxy](https://github.com/galaxyproject/ansible-galaxy)

# Job Execution on External Resources

Pulsar is a distributed job execution application built for Galaxy: [github.com/galaxyproject/pulsar](https://github.com/galaxyproject/pulsar)

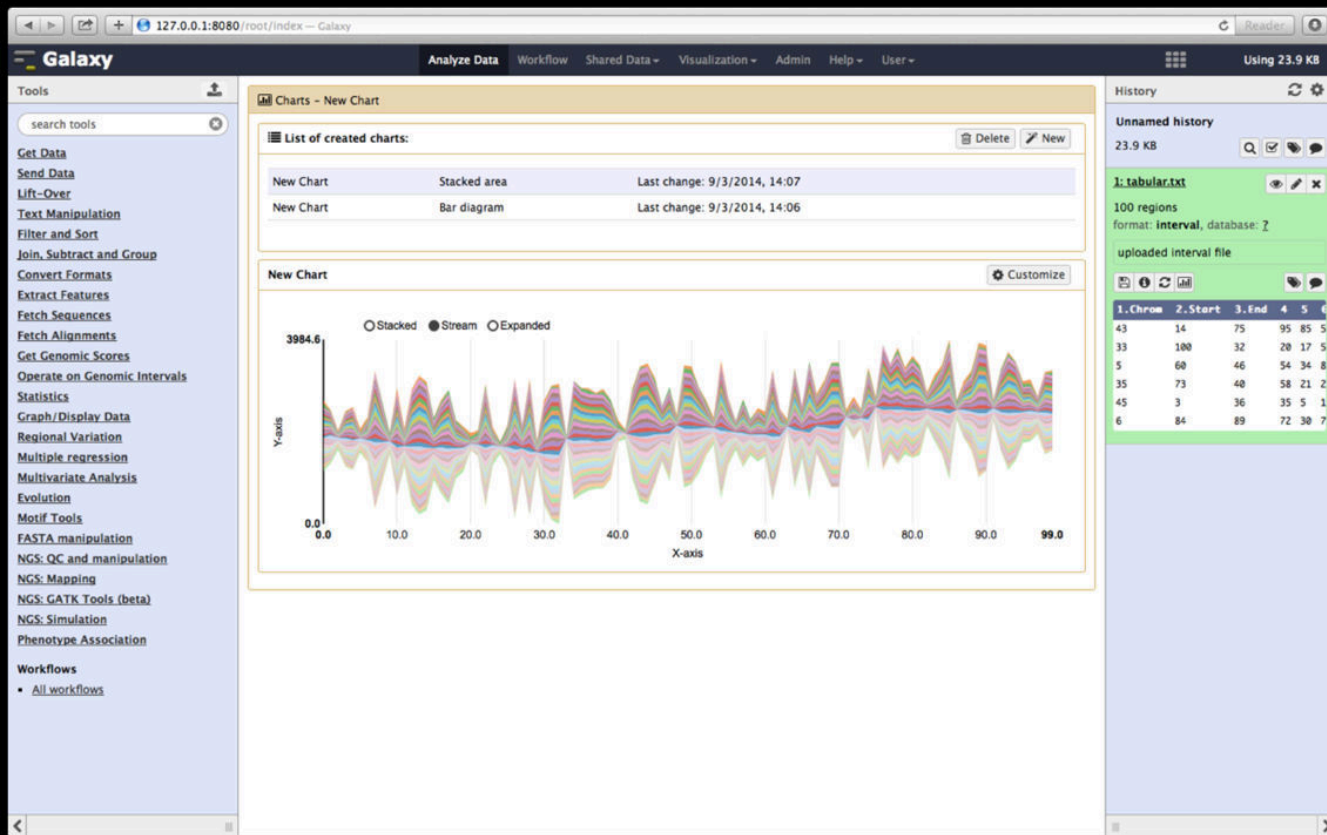
\*Enabled On Main\* for particular tools (for now): BWA, Bowtie, Bowtie2, Tophat, Tophat2, Lastz, Megablast, Cuff\* with double the cores and permitted walltime per job.

## XSEDE Allocation



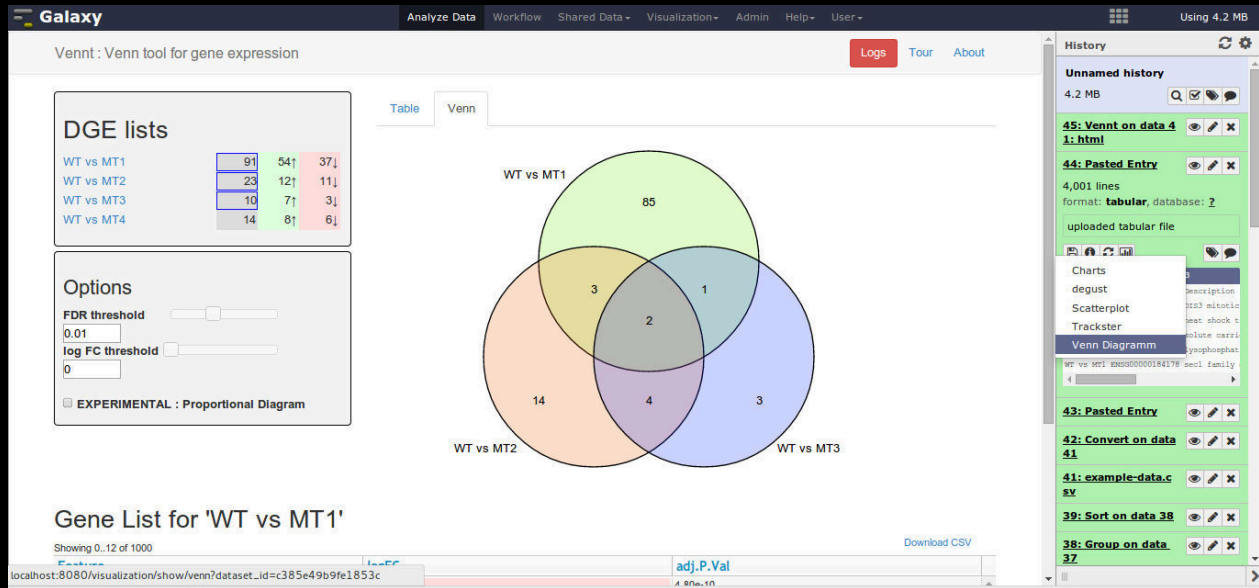
The screenshot shows a web form titled "Job Resource Parameters:". It contains a dropdown menu with the text "Specify job resource parameters". Below this is a section titled "Compute Resource:" with a dropdown menu showing "TACC Stampede (beta)". A link below the dropdown reads "Need help selecting a compute resource? Options and limits are explained in detail [in the Galaxy Wiki](#)". At the bottom of the form is a blue button labeled "Execute".

# Viz Framework





# Viz Framework



Simon Gladman, Björn Grüning <https://github.com/Slugger70/Vennt-galaxy-vis/>

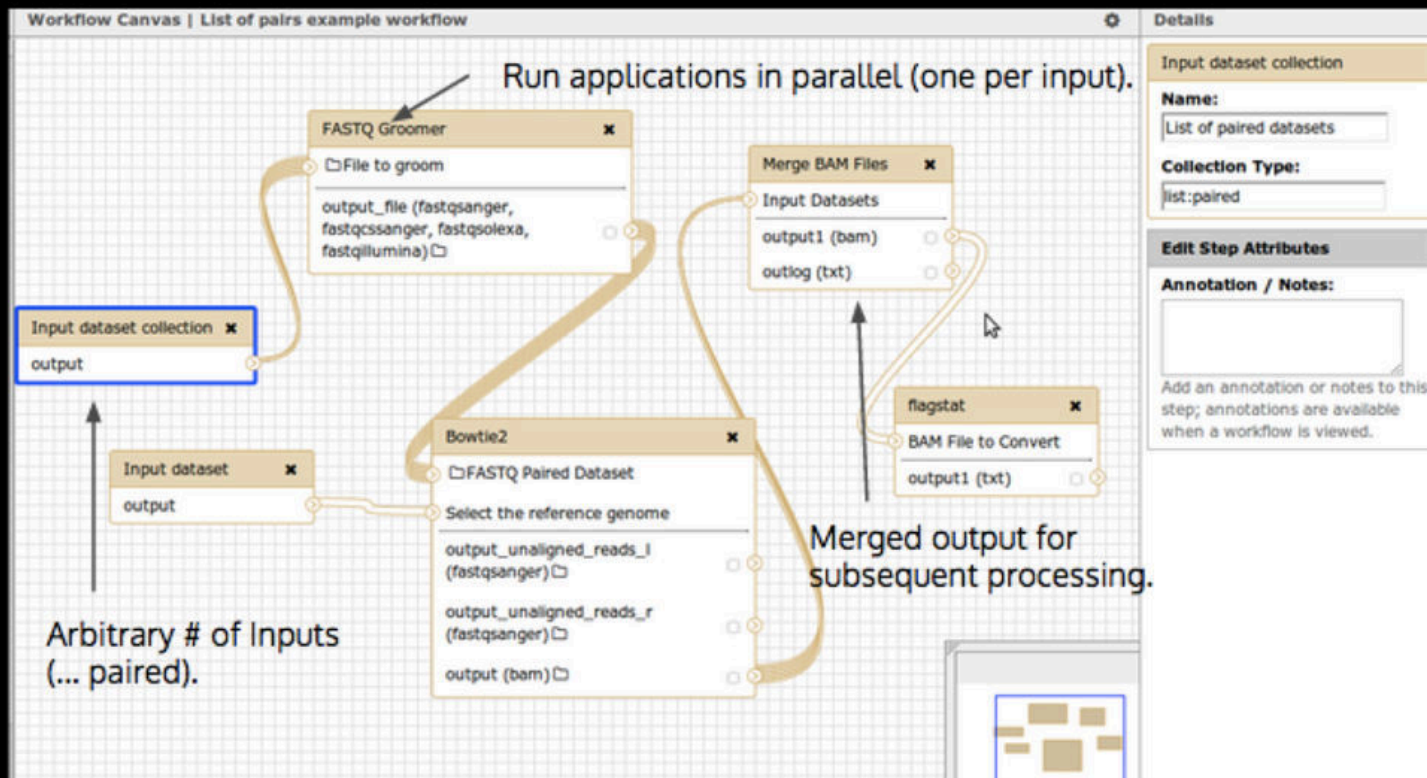
# Data Libraries and other UI Evolution

It's all about the API.

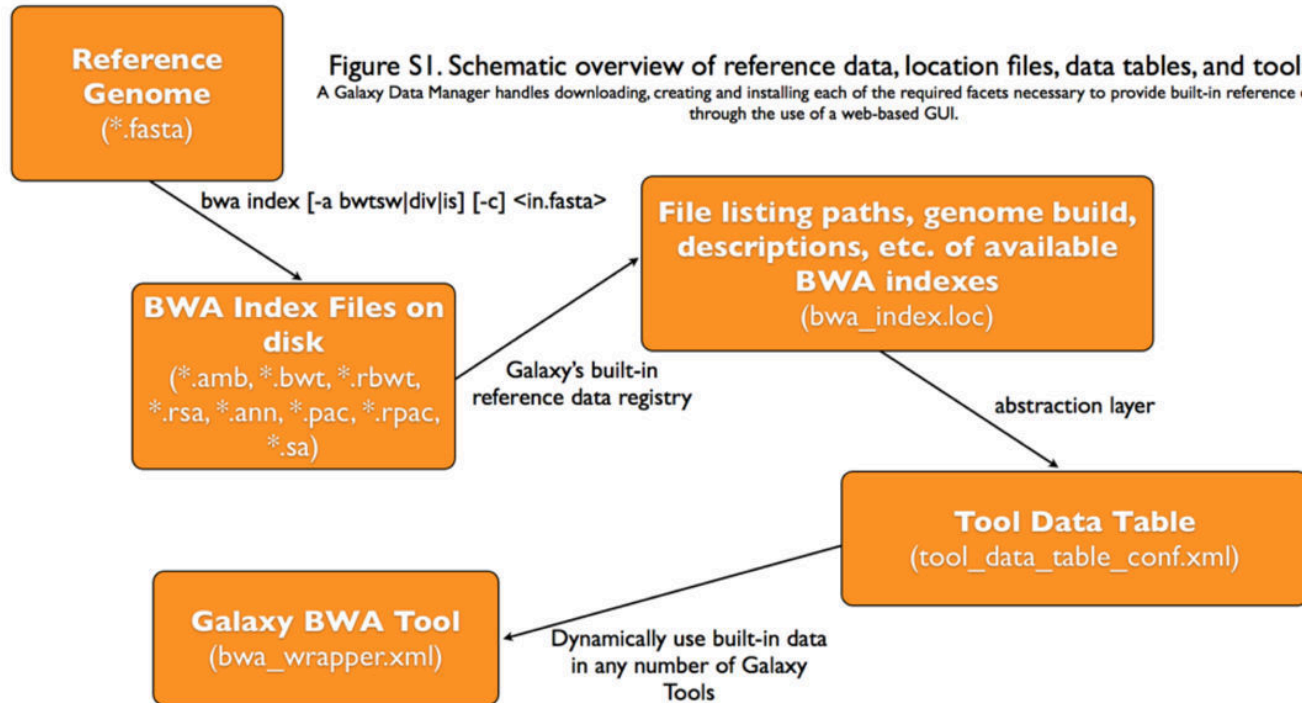
Shared Data -> “Data Libraries Beta” -- it's FAST!

Histories (filters, multiple dataset operations, multiple history views), Tool Rendering, and more.

# Dataset Collections



# Data Managers



# Community Interface / Help Desk

The screenshot displays the Biostars website interface, which serves as a community hub and help desk for Galaxy users. The page is titled "Latest Posts" and shows the URL <https://biostar.usegalaxy.org>. The navigation bar includes links for "LATEST", "OPEN", "RNA-SEQ", "CHIP-SEQ", "SNP", "ASSEMBLY", "FORUM", "PLANET", and "ALL". A welcome message reads "Welcome to Galaxy Biostar! User support for Galaxy!" with links to "about", "faq", and "rss". The main content area features a search bar with "Live search: start typing..." and a "Classic search" option. Below the search bar, a list of posts is displayed, each with a title, vote count, answer count, view count, tags, and author information. The posts include:

- combining indexed paired reads from two lanes of a Hiseq rapidrun** (0 votes, 0 answers, 4 views, tags: ma-seq, written 8 hours ago by mduffy)
- Forum: Need help with "Tophat2" tool** (0 votes, 0 answers, 4 views, tags: tophat2, forum, written 12 hours ago by martine.collart)
- WIG file from MACS2** (0 votes, 1 answer, 7 views, tags: macs2, cistrome, written 16 hours ago by katherinedixon4)
- I want to delete my account** (0 votes, 1 answer, 10 views, tag: galaxy, written 1 day ago by christine.noite)
- phyloP not recognizing any of the HG19 tags as being HG19** (0 votes, 1 answer, 20 views, tags: metadata, phyloP, written 3 days ago by caffetazbrant)
- Problems adding build-in genomes for tophat on local galaxy** (1 vote, 2 answers, 28 views, tags: galaxy, tophat, written 4 days ago by bimbam42)
- rsID from list of chromosome and position** (0 votes, 1 answer, 15 views, tag: snp, written 2 days ago by sookwah.yee)

On the right side, there are sections for "Recent Votes" and "Recent Locations". The "Recent Votes" section lists votes on various topics, including "Problems adding build-in genomes for tophat on local galaxy" and "Need help with 'Variant Annotator' tool". The "Recent Locations" section shows the locations of users who visited the site, such as "United Kingdom", "United States", "Netherlands", and "European Union". At the bottom left, a traffic report indicates "15 users visited in the last hour".

# Toolshed

(particularly dependency management)

```
<actions>
  <action type="shell_command">git clone --recursive git://github.com/ekg/freebayes.git</action>
  <action type="shell_command">git checkout 9608597d12e127c847ae03aa03440ab63992fedf</action>
  <action type="shell_command">git submodule update --recursive</action>
  <action type="shell_command">make || ( make clean &&; sed -i.bak -e &apos;s:LIBS = -lz -lm -L./ -L../vcflib/tabixpp/
  <action type="move_directory_files">
    <source_directory>bin</source_directory>
    <destination_directory>$INSTALL_DIR/bin</destination_directory>
  </action>
```

# Ongoing Efforts

(see trello for details!)

# Workflow Engine Overhaul

- 1) Stateful execution model
  - enables control structures
- 2) Make it pluggable



# Niche-specific Tools, Guides, Workflows

Highly curated guides for particular analysis approaches - more than just Pages

Integrated 'tour' of Galaxy analysis

# Federation / Distributed Data

\*pretty much everything\* should be portable

Especially important due to growing numbers of separate public, cloud, and local instances.

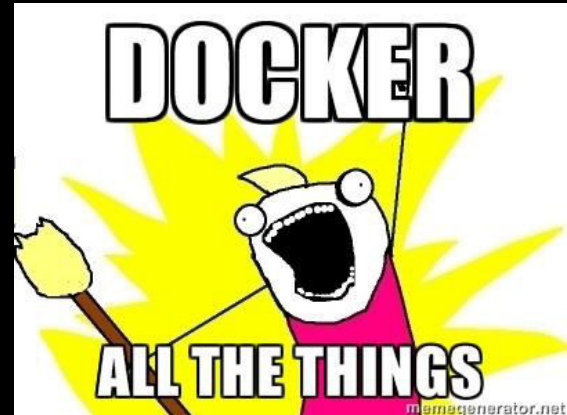
# ToolShed, Tool Development Evolution

GitHub-backed repositories for more developer engagement

Submission process overhaul

Major UI overhaul

Tooling for local developers



# Interface for truly large-scale analysis

The 'history' concept is fantastic for small exploratory analysis.

It does not scale for people needing to run thousands of samples.

# Scalable Training

Cloud 'workshop in a box' and similar initiatives to distribute training materials and knowledge amongst the community.

# Sample Tracking

App - based handling of samples

# Thanks!

Keep the conversation going:

[chat.freenode.net #galaxyproject](https://chat.freenode.net/#galaxyproject)

[galaxy-dev@lists.bx.psu.edu](mailto:galaxy-dev@lists.bx.psu.edu)

<http://galaxyproject.org/trello>

[https://wiki.galaxyproject.org/GalaxyUpdates/2014\\_10#Galaxy\\_Needs\\_Your\\_Input.21](https://wiki.galaxyproject.org/GalaxyUpdates/2014_10#Galaxy_Needs_Your_Input.21)