Galaxy Project Introduction

Galaxy Demo / Training Day
September 30, 2014, Bern

Hans-Rudolf Hotz (hrh@fmi.ch)

Friedrich Miescher Institute for Biomedical Research
Basel, Switzerland

the Galaxy Project

Galaxy

FMI

Friedrich Miescher Institute
for Biomedical Research

the Galaxy Project

Galaxy is an open,
web-based platform for
accessible, *reproducible*, and *transparent*
computational biomedical research.

http://galaxyproject.org/

Friedrich Miescher Institute
for Biomedical Research

the Galaxy Project

Galaxy

Accessible: Users without programming experience can easily specify parameters and run tools and workflows.

Reproducible: Galaxy captures information so that any user can repeat and understand a complete computational analysis.

Transparent: Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

http://galaxyproject.org/

the Galaxy Project

Galaxy

Galaxy is *an open*,
web-based platform for
accessible, reproducible, and transparent
computational biomedical research.

http://galaxyproject.org/

FMI
Friedrich Miescher Institute
for Biomedical Research

the Galaxy Project

Galaxy

Galaxy is available as

a free web server

an open source software

FMI
Friedrich Miescher Institute
for Biomedical Research

the Galaxy Project

Galaxy is available as

a free web server

an open source software

These options result in several ways to use Galaxy

Friedrich Miescher Institute for Biomedical Research

Galaxy is available as

Galaxy

As a free web server integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage

https://usegalaxy.org/

Galaxy is available as

Galaxy

As a free web server integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage

https://usegalaxy.org/

a centralized solution cannot support the different analysis needs of the entire world.

FMI
Friedrich Miescher Institute
for Biomedical Research

Galaxy is available as

Galaxy

As an open source software

https://getgalaxy.org/

FMI
Friedrich Miescher Institute
for Biomedical Research

Galaxy is available as

Galaxy

As an open source software

https://getgalaxy.org/

...and it is installed in
locations around the world

FMI
Friedrich Miescher Institute
for Biomedical Research

deploy Galaxy

Galaxy

install Galaxy locally

use Galaxy on the Cloud

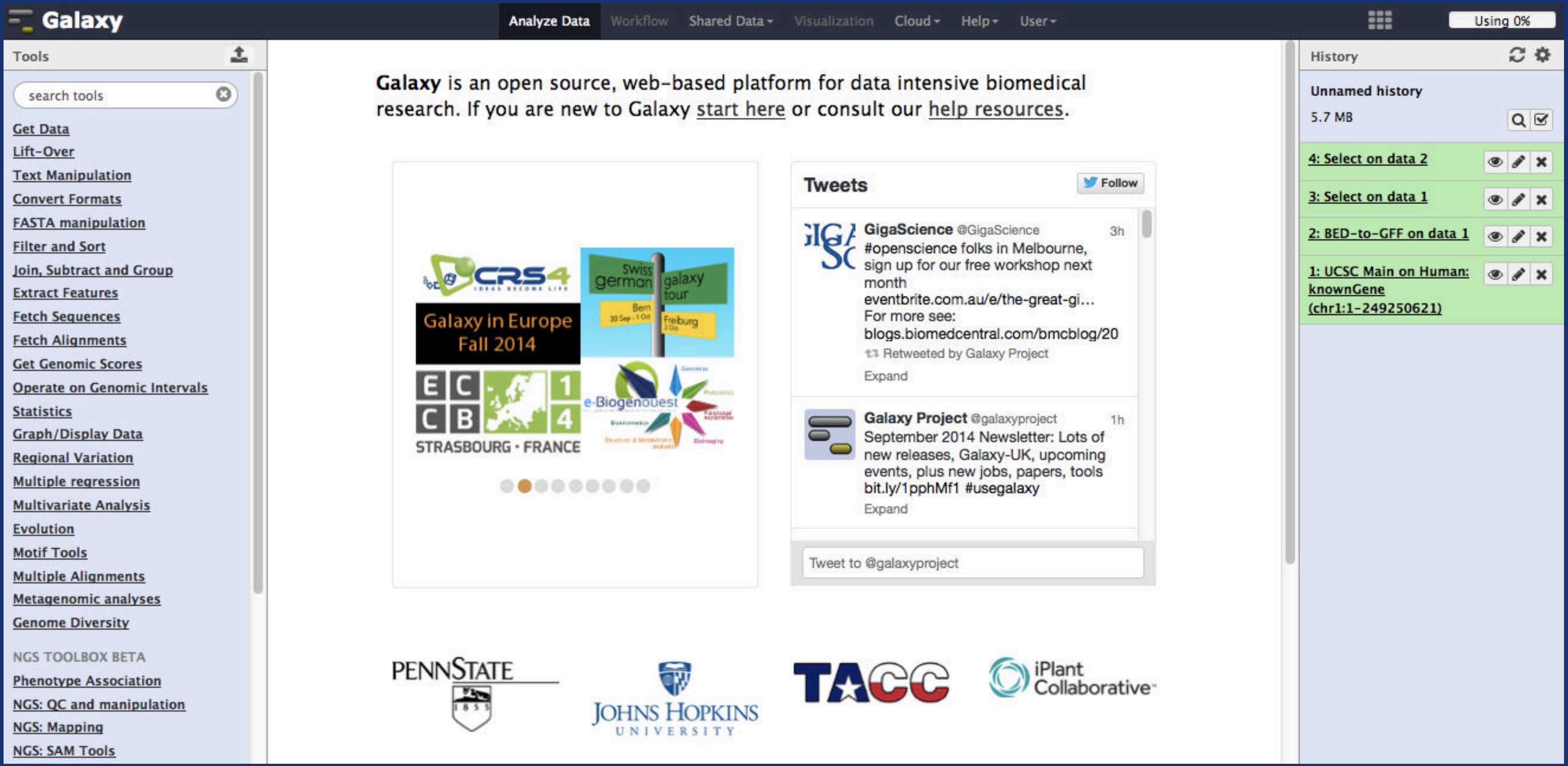
get the SlipStream Galaxy appliance (BioTeam)

FMI
Friedrich Miescher Institute
for Biomedical Research

using Galaxy

Galaxy

Galaxy's public web site



Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources.

GigaScience @GigaScience 3h
#openscience folks in Melbourne, sign up for our free workshop next month eventbrite.com.au/e/the-great-gi… For more see: blogs.biomedcentral.com/bmcblog/20
Retweeted by Galaxy Project

Galaxy Project @galaxyproject 1h
September 2014 Newsletter: Lots of new releases, Galaxy-UK, upcoming events, plus new jobs, papers, tools bit.ly/1pphMf1 #usegalaxy

http://usegalaxy.org

FMI
Friedrich Miescher Institute
for Biomedical Research

learning Galaxy

Galaxy

http://wiki.galaxyproject.org/Learn

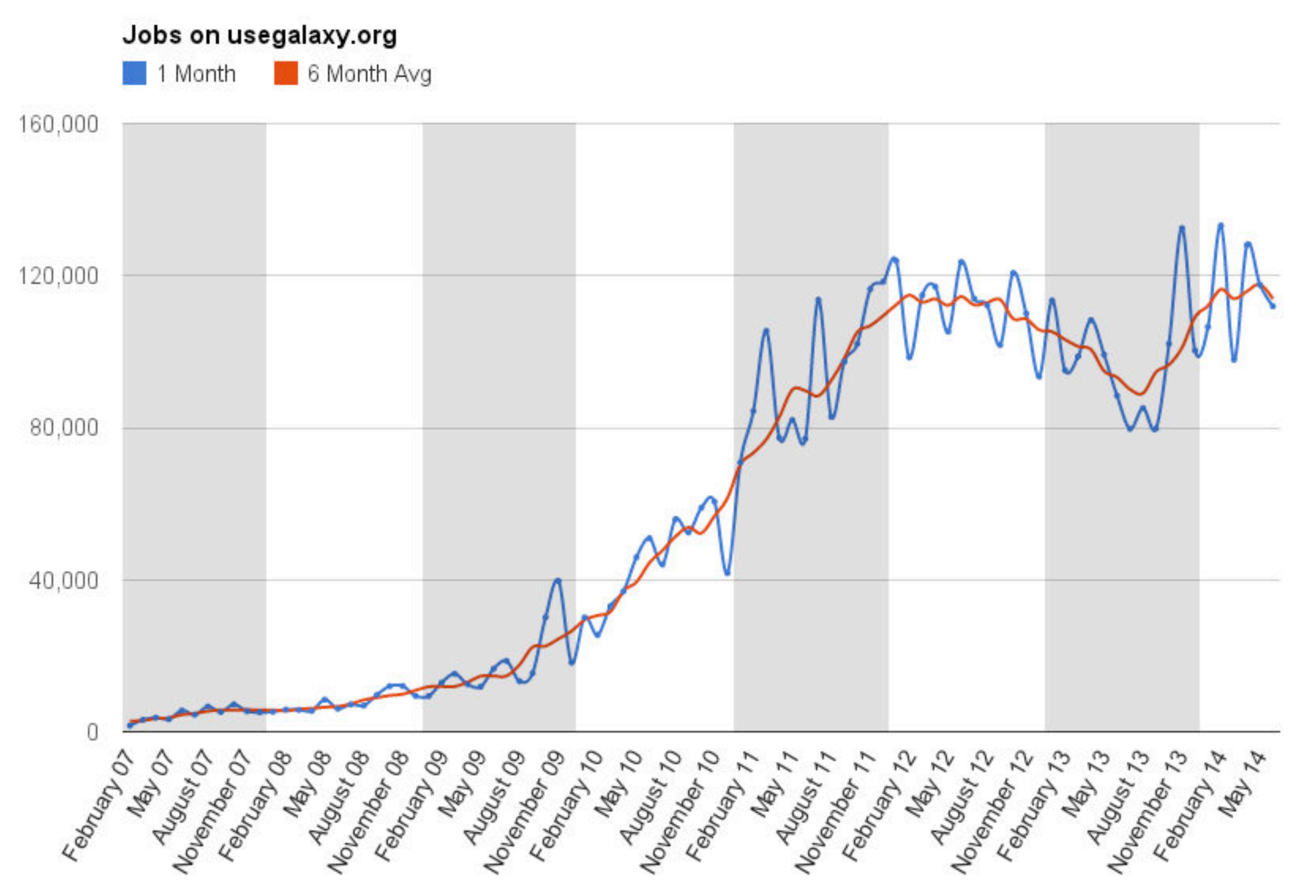
tutorials
screencats
shared histories & workflows

http://wiki.galaxyproject.org/CitingGalaxy

http://wiki.galaxyproject.org/MailingLists

FMI
Friedrich Miescher Institute
for Biomedical Research

and Galaxy is used



Friedrich Miescher Institute
for Biomedical Research

Galaxy is a community

Galaxy

FMI

Friedrich Miescher Institute
for Biomedical Research

Galaxy is a community

Galaxy

in addition to *http://usegalaxy.org* there are 60 public instances offered by institutes from all over the world

Galaxy is a community

Galaxy

in addition to *http://usegalaxy.org* there are 60 public instances offered by institutes from all over the world

plus an unknown number of private Galaxy servers

FMI

Friedrich Miescher Institute
for Biomedical Research

Galaxy is a community

Galaxy

in addition to *http://usegalaxy.org* there are 60 public instances offered by institutes from all over the world

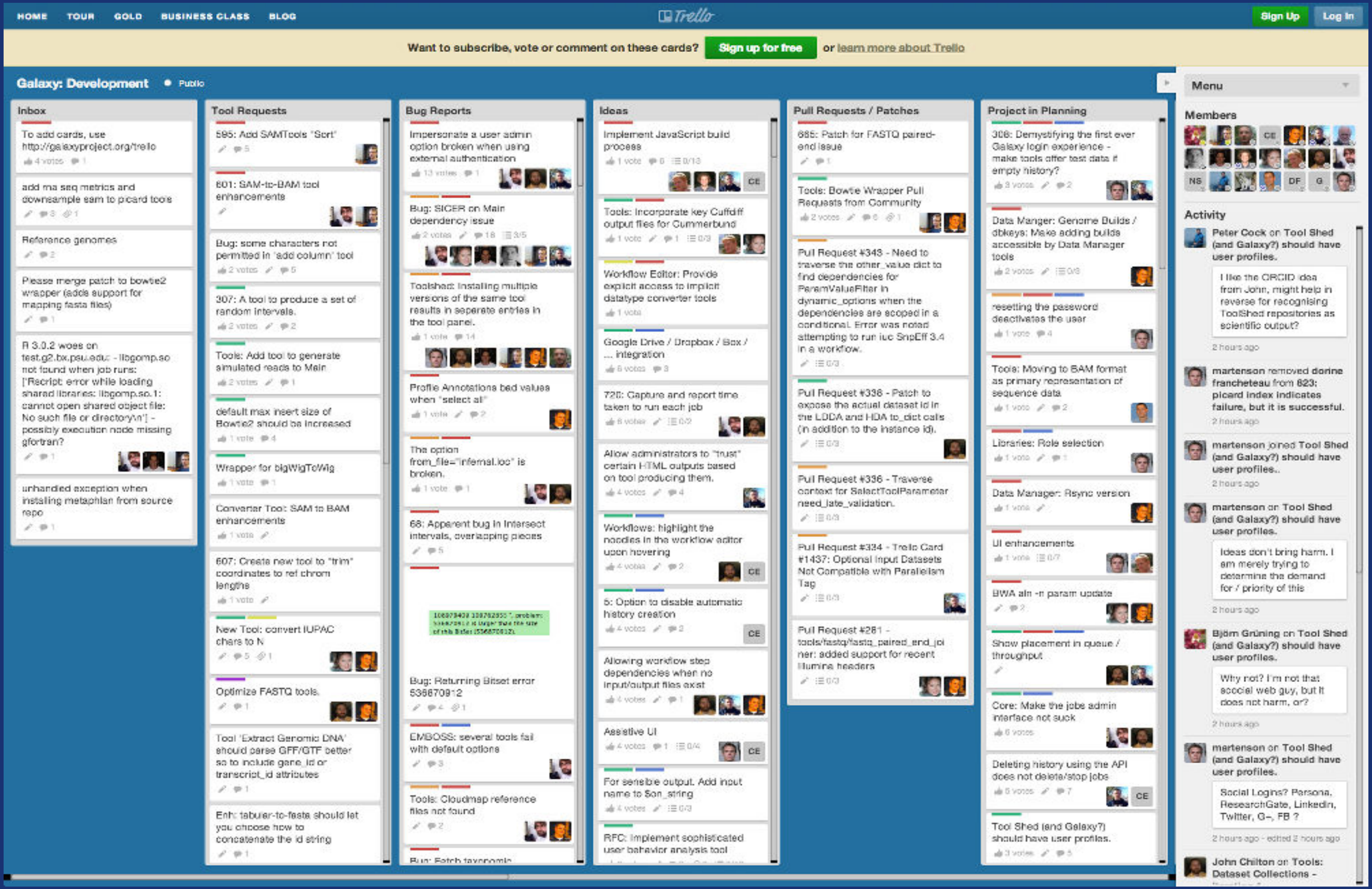
plus an unknown number of private Galaxy servers

~4000 code commits from over 50 people per year

FMI

Friedrich Miescher Institute
for Biomedical Research

Galaxy is a community



Want to subscribe, vote or comment on these cards? Sign up for free or learn more about Trello

Galaxy: Development

Inbox
- To add cards, use http://galaxyproject.org/trello (4 votes)
- add rna seq metrics and downsample sam to picard tools
- Reference genomes
- Please merge patch to bowtie2 wrapper (adds support for mapping fasta files)
- R 3.0.2 woes on test.g2.bx.psu.edu - libgomp.so not found when job runs: ['Rscript error while loading shared libraries: libgomp.so.1: cannot open shared object file: No such file or directory\n'] - possibly execution node missing gfortran?
- unhandled exception when installing metaphlan from source repo

Tool Requests
- 595: Add SAMTools "Sort"
- 601: SAM-to-BAM tool enhancements
- Bug: some characters not permitted in 'add column' tool (2 votes)
- 307: A tool to produce a set of random intervals. (2 votes)
- Tools: Add tool to generate simulated reads to Main (2 votes)
- default max insert size of Bowtie2 should be increased (1 vote)
- Wrapper for bigWigToWig (1 vote)
- Converter Tool: SAM to BAM enhancements (1 vote)
- 607: Create new tool to "trim" coordinates to ref chrom lengths (1 vote)
- New Tool: convert IUPAC chars to N
- Optimize FASTQ tools.
- Tool 'Extract Genomic DNA' should parse GFF/GTF better so to include gene_id or transcript_id attributes
- Enh: tabular-to-fasta should let you choose how to concatenate the id string

Bug Reports
- Impersonate a user admin option broken when using external authentication (13 votes)
- Bug: SICER on Main dependency issue (2 votes)
- Toolshed: Installing multiple versions of the same tool results in seperate entries in the tool panel. (1 vote)
- Profile Annotations bad values when "select all" (1 vote)
- The option from_file="infernal.loc" is broken. (1 vote)
- 68: Apparent bug in Intersect intervals, overlapping pieces
- Bug: Returning Bitset error 536870912
- EMBOSS: several tools fail with default options
- Tools: Cloudmap reference files not found
- Bug: Fetch taxonomic

Ideas
- Implement JavaScript build process (1 vote)
- Tools: Incorporate key Cuffdiff output files for Cummerbund (1 vote)
- Workflow Editor: Provide explicit access to implicit datatype converter tools (1 vote)
- Google Drive / Dropbox / Box / ... integration (6 votes)
- 720: Capture and report time taken to run each job (6 votes)
- Allow administrators to "trust" certain HTML outputs based on tool producing them. (4 votes)
- Workflows: highlight the noodles in the workflow editor upon hovering (4 votes)
- 5: Option to disable automatic history creation (4 votes)
- Allowing workflow step dependencies when no input/output files exist (4 votes)
- Assistive UI (4 votes)
- For sensible output. Add input name to $on_string (4 votes)
- RFC: Implement sophisticated user behavior analysis tool

Pull Requests / Patches
- 685: Patch for FASTQ paired-end issue
- Tools: Bowtie Wrapper Pull Requests from Community (2 votes)
- Pull Request #343 - Need to traverse the other_value dict to find dependencies for ParamValueFilter in dynamic_options when the dependencies are scoped in a conditional. Error was noted attempting to run iuc SnpEff 3.4 in a workflow.
- Pull Request #338 - Patch to expose the actual dataset id in the LDDA and HDA to_dict calls (in addition to the instance id).
- Pull Request #336 - Traverse context for SelectToolParameter need_late_validation.
- Pull Request #334 - Trello Card #1437: Optional Input Datasets Not Compatible with Parallelism Tag
- Pull Request #281 - tools/fastq/fastq_paired_end_joiner: added support for recent Illumina headers

Project in Planning
- 308: Demystifying the first ever Galaxy login experience - make tools offer test data if empty history? (3 votes)
- Data Manger: Genome Builds / dbkeys: Make adding builds accessible by Data Manager tools (2 votes)
- resetting the password deactivates the user (1 vote)
- Tools: Moving to BAM format as primary representation of sequence data (1 vote)
- Libraries: Role selection (1 vote)
- Data Manager: Rsync version (1 vote)
- UI enhancements (1 vote)
- BWA aln -n param update
- Show placement in queue / throughput
- Core: Make the jobs admin interface not suck (6 votes)
- Deleting history using the API does not delete/stop jobs (5 votes)
- Tool Shed (and Galaxy?) should have user profiles. (3 votes)

Members

CE NS DF G

Activity

Peter Cock on **Tool Shed (and Galaxy?) should have user profiles.**

I like the ORCID idea from John, might help in reverse for recognising ToolShed repositories as scientific output?

2 hours ago

martenson removed **dorine francheteau** from **623: picard index indicates failure, but it is successful.**

2 hours ago

martenson joined **Tool Shed (and Galaxy?) should have user profiles..**

2 hours ago

martenson on **Tool Shed (and Galaxy?) should have user profiles.**

Ideas don't bring harm. I am merely trying to determine the demand for / priority of this

2 hours ago

Björn Grüning on **Tool Shed (and Galaxy?) should have user profiles.**

Why not? I'm not that soccial web guy, but it does not harm, or?

2 hours ago

martenson on **Tool Shed (and Galaxy?) should have user profiles.**

Social Logins? Persona, ResearchGate, Linkedin, Twitter, G+, FB ?

2 hours ago - edited 2 hours ago

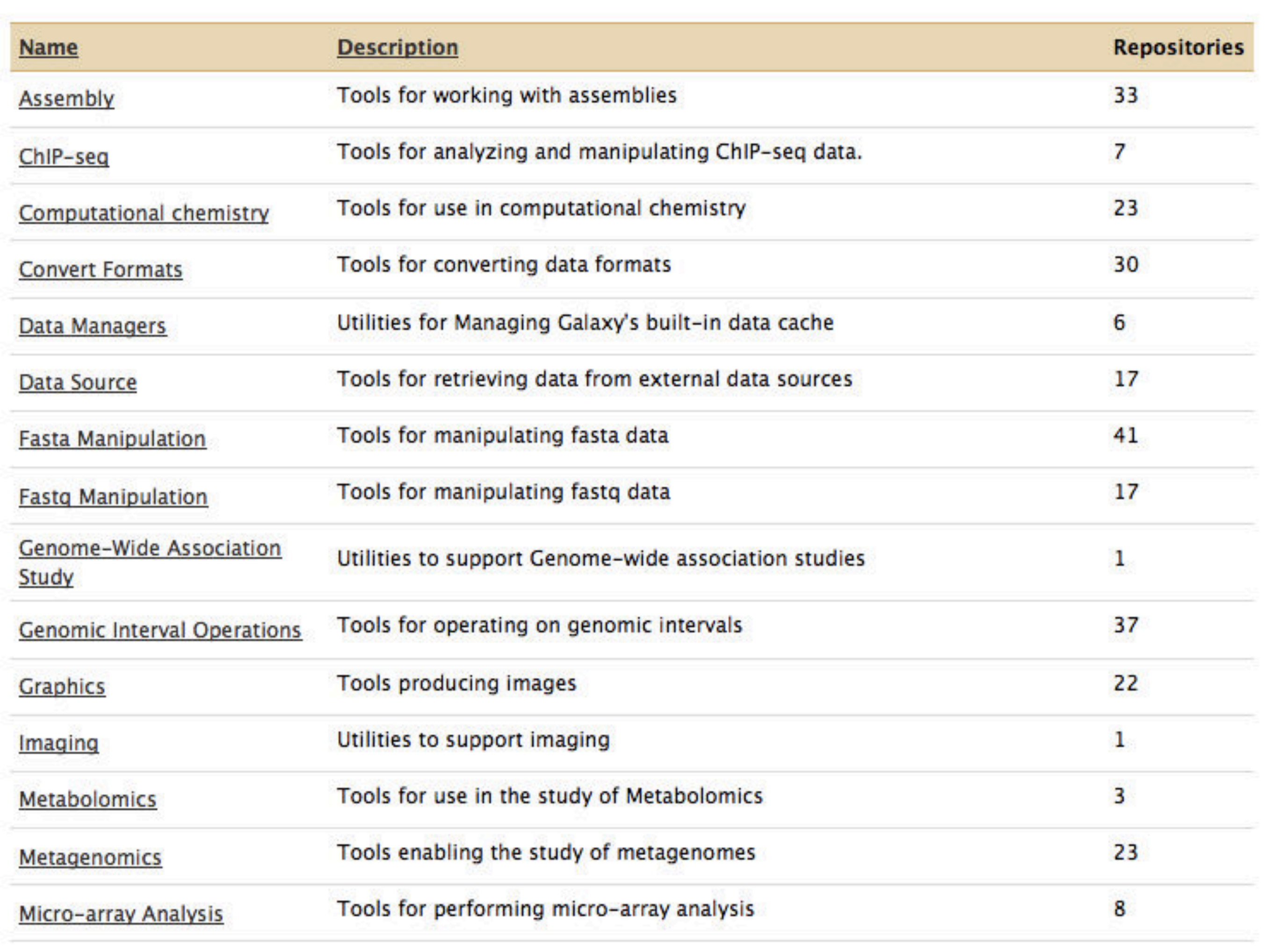
John Chilton on **Tools: Dataset Collections -**

http://bit.ly/gxytrello

Galaxy toolshed

Galaxy

2570 'tools' published from ~220 developers



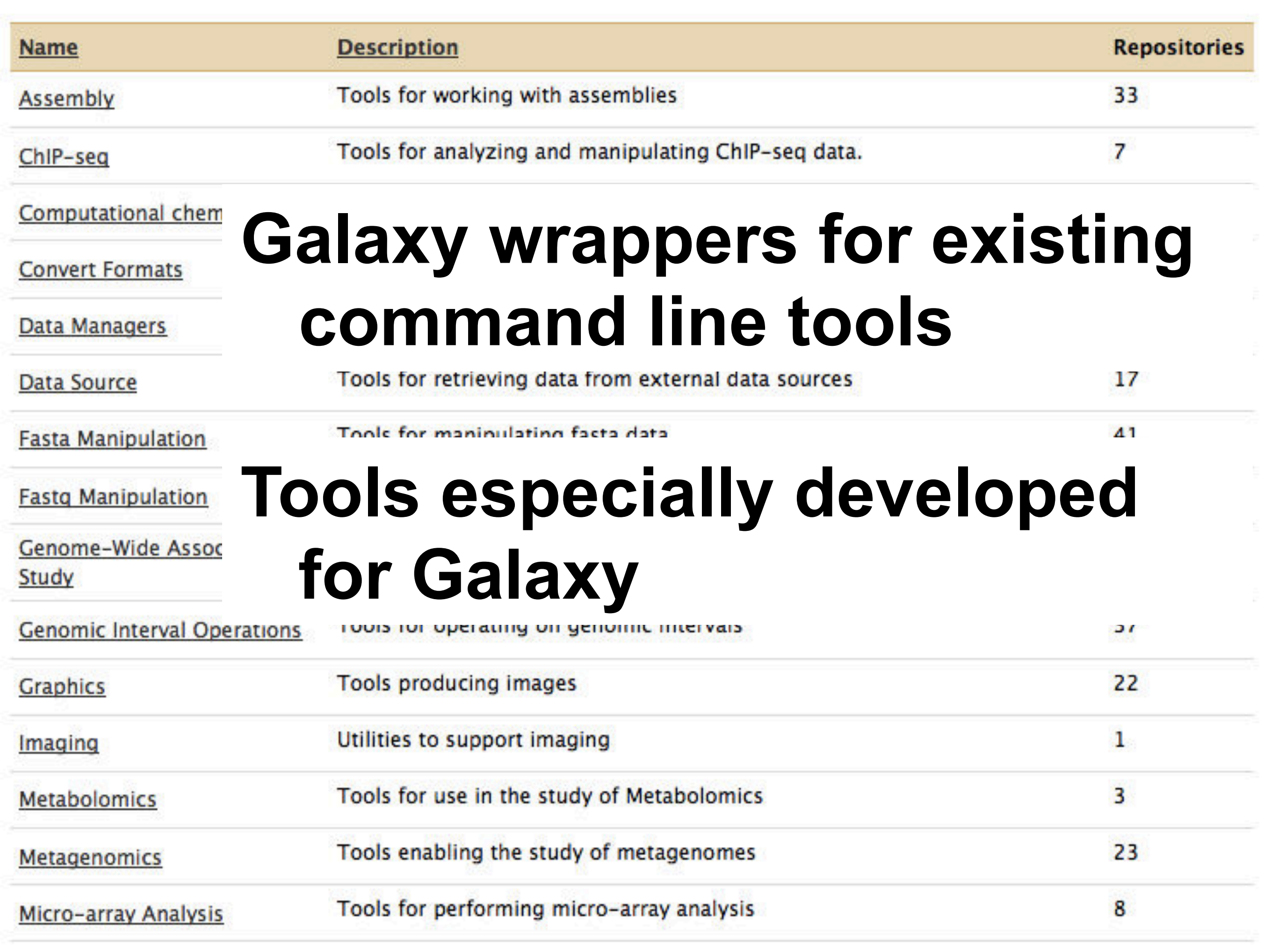
Name	Description	Repositories
Assembly	Tools for working with assemblies	33
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	7
Computational chemistry	Tools for use in computational chemistry	23
Convert Formats	Tools for converting data formats	30
Data Managers	Utilities for Managing Galaxy's built-in data cache	6
Data Source	Tools for retrieving data from external data sources	17
Fasta Manipulation	Tools for manipulating fasta data	41
Fastq Manipulation	Tools for manipulating fastq data	17
Genome-Wide Association Study	Utilities to support Genome-wide association studies	1
Genomic Interval Operations	Tools for operating on genomic intervals	37
Graphics	Tools producing images	22
Imaging	Utilities to support imaging	1
Metabolomics	Tools for use in the study of Metabolomics	3
Metagenomics	Tools enabling the study of metagenomes	23
Micro-array Analysis	Tools for performing micro-array analysis	8

FMI

Friedrich Miescher Institute
for Biomedical Research

Galaxy toolshed

2570 'tools' published from ~220 developers



Name	Description	Repositories
Assembly	Tools for working with assemblies	33
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	7
Computational chem		
Convert Formats		
Data Managers		
Data Source	Tools for retrieving data from external data sources	17
Fasta Manipulation	Tools for manipulating fasta data	41
Fastq Manipulation		
Genome-Wide Assoc Study		
Genomic Interval Operations	Tools for operating on genomic intervals	37
Graphics	Tools producing images	22
Imaging	Utilities to support imaging	1
Metabolomics	Tools for use in the study of Metabolomics	3
Metagenomics	Tools enabling the study of metagenomes	23
Micro-array Analysis	Tools for performing micro-array analysis	8

Galaxy wrappers for existing command line tools

Tools especially developed for Galaxy

FMI
Friedrich Miescher Institute
for Biomedical Research

The Galaxy Team

Galaxy

Penn State

Johns Hopkins

https://wiki.galaxyproject.org/GalaxyTeam

GCC 2015

Galaxy Community Conference

6-8th July 2015

The Sainsbury Laboratory
Norwich, UK

galaxyproject.org

let's do some work now

Galaxy

http://cloud1.galaxyproject.org

FMI

Friedrich Miescher Institute
for Biomedical Research