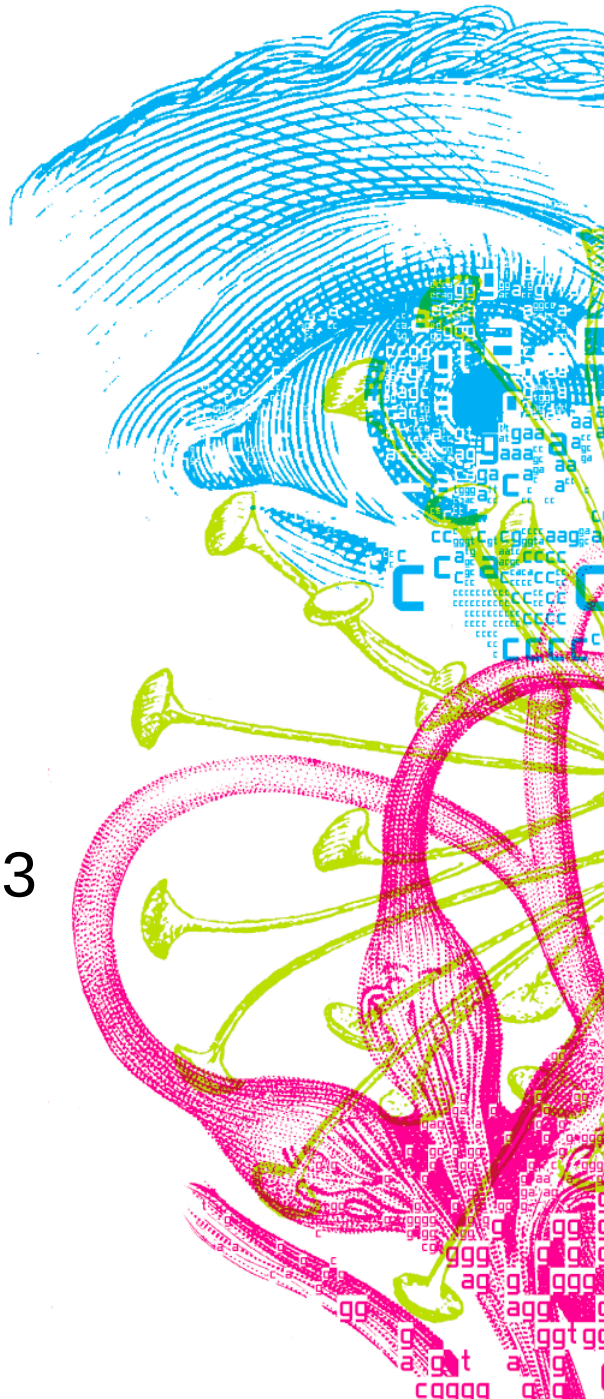
Netherlands Bioinformatics Centre

NGS task force & GCC2013 report

http://wiki.galaxyproject.org/Events/GCC2013



Update

- NGS PI meeting on June 18th
  - 2014 onwards, monthly meetings or thematic meetings?
  - SV benchmarking
  - Cross sectors, e.g. PigVD (based on DVD)
  - NGS course portfolio
- Galaxy community conference, June 30th-July 2nd, Oslo
  - RNAseq training, ~60 attendees
  - Amazon performance is disappointing

GCC 2013

- ~210 attendees
  - NBIC: Leon
  - LUMC: Jeroen, Bowo, Wai Yi
  - EMC: Saskia, Rene
  - NIZO: Judith
  - UvA/NLeSC: Mateusz
  - LU:
  - WUR: Eric, Pieter

- Single track 20 presentations, ~30 posters, and many lightening talks!
  - People are sober thanks to the 5 euro beers.
  - All slides/videos on line
    http://wiki.galaxyproject.org/Events/GCC2013/Program

nbic Netherlands Bioinformatics Centre

Updates from Galaxy team

- UI improvement to handle large # datasets
  - avoid over-blowing histories
- Integrate toolshed & data manager
  - Install all tools, dependencies, built-in data via Admin panel
  - Talks from Greg Von Kuster and Daniel Blankenberg
- Galaxy to become a generic platform, no tools/data associated with the vanilla version.
  - Several scripts will be created to populate the vanilla Galaxy with standard NGS tools, data.
- Toolshed will be contributed by 3rd party developers and monitored by the IUC (Intergalactic Utilities Commission)
  - Talk from Dannon Baker

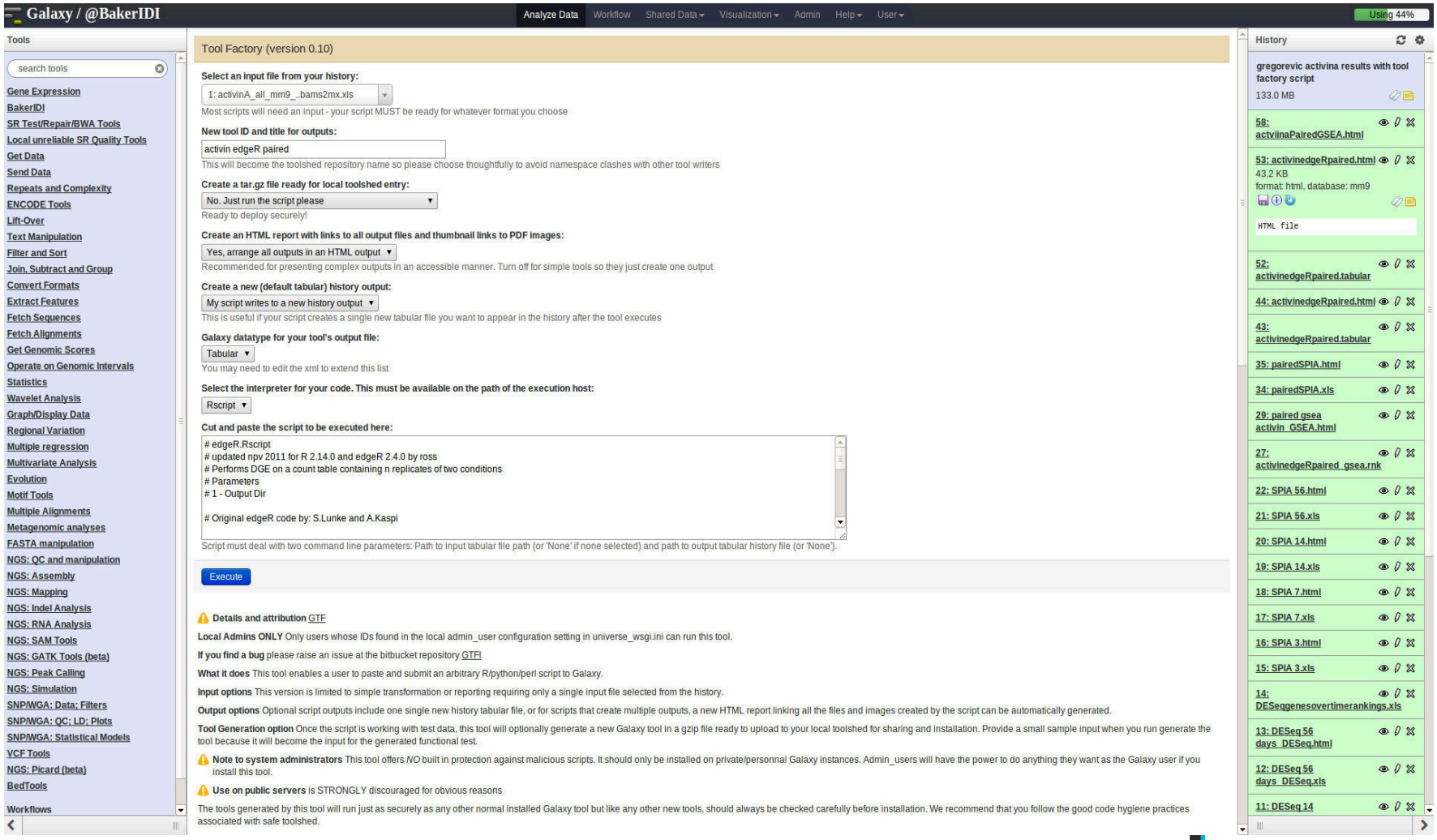
nbic Netherlands Bioinformatics Centre

Shared Interest

- Reproducibility
  - Versioning tools, keep histories, use test, etc.
- Cloud!!!
  - Amazon
  - EMC
  - Other private clouds (in Germany, France, US, Norway, etc), most OpenStack based.
- Professionalization and business model of Galaxy
  - BioTeam SlipStream
- Non-NGS
  - Galaxy-P, cheminformatics, image processing

nbic Netherlands Bioinformatics Centre

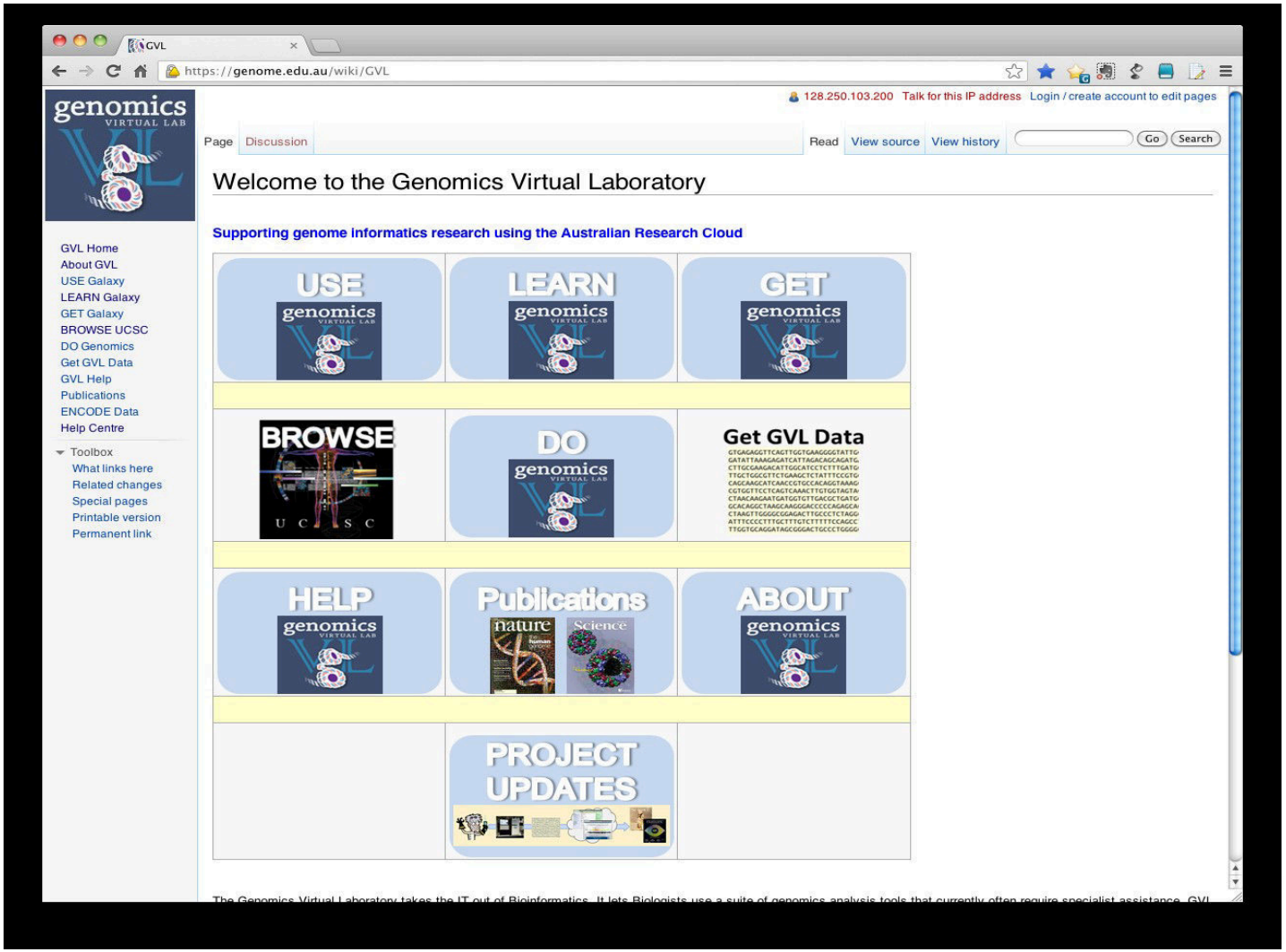
Highlight #1 (Tool factory, Ross Lazarus)



Netherlands Bioinformatics Centre

Highlight #2 (Galaxy tutorial)

https://genome.edu.au/wiki/Galaxy_Tutorials

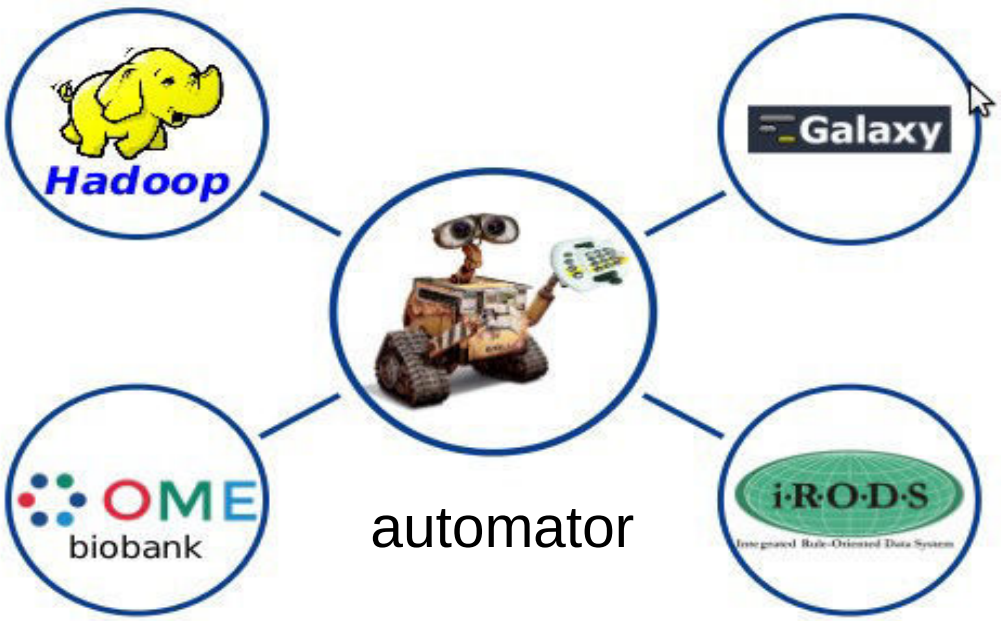


Netherlands Bioinformatics Centre

Highlight #3 Automated processing and tracking platform

Luca Pireddu, CRS4

- Seal: toolkit for Hadoop-based sequencing data processing
  - demultiplexing, alignment (based on BWA, sorting, etc.)
- Pydoop: Python API for Hadoop
  - A dependency for Seal, but also used for custom tools and scripts
- SeqPig: SQL-like scripting for Hadoop with sequencing-specic functionality



iRODS, distributed file management system, including optimal file transfer support.

nbic Netherlands Bioinformatics Centre

Highlight #4 Genomics Hyperbrowser & Gtrack data type

- Hyperbrowser
  - Geir K Sandve
  - Statistical analysis tool for genomics tracks
  - http://hyperbrowser.uio.no/test/ Including tutorialss

- Gtrack
  - Sveinung Gundersen
  - A new datatype to harmonize the existing datatypes
  - general purpose, tabular file format for representing data in the form of genomic tracks
  - Several tools and converters available

Highlight #5 Auditing Galaxy for clinical use

- Sanjay Joshi, from EMC
  - The Clinical Galaxy: A validated platform initiative "We will present an overview of the requirements to move Galaxy into the Clinical realm."

Highlight #6 BioBlend - automating bioinformatics with Galaxy and CloudMan

- Clare Sloggett
  - http://bioblend.readthedocs.org/en/latest/
  - https://github.com/afgane/bioblend/

Contributed by Mateusz Kuzak

nbic Netherlands Bioinformatics Centre

Public server BOF group discussion

- Security and Billing
  - Galaxy is not designed with security in mind from the ground up.
  - Authentication needs to be more pluggable.
  - Galaxy lacks the reporting feature on cpu hours which is sometimes very useful for funding agency.
- Releases
  - Galaxy lacks a stable release scheme (~twice per year) which makes the life of public Galaxy admin a lot easier.
  - Better Versioning
- Tool shed is currently making things more complicated.
- Dataset profiligation. Can easily end up with 3 copies of most of your datasets, just to get files into Galaxy.
- maintaining a public Galaxy server well requires minimal 0.5 fte

nbic Netherlands Bioinformatics Centre