usegalaxy.

Global Galaxy for Everyone

Simon Gladman, 2018
& Bjoern and Nate
Public Galaxy servers - In the beginning..

usegalaxy.org - Galaxy Main

- Run by many institutions in the USA
  - Penn State, Johns Hopkins, Oregon Health and Science Uni.
  - TACC, JetStream, CyVerse, Massive
- Many users
- Constantly large queue
Public Galaxy servers - And the list grew..

- Different communities
- Different Galaxy versions
- Different tools
- Different references
- Different resources
- Duplication of effort
So some of us got together..

- At Galaxy Australasia Meeting 2017
  - Some interested people from US, Europe and Australia
  - Agreed to work together to support publically available Galaxys
  - Sharing software/reference resources
  - Do as much together as possible

And so usegalaxy.* was born
What is usegalaxy.*?

Group of public Galaxy servers

- Present a similar experience to users no matter which they use
- Guarantee a minimum service
  - Tools & versions
  - Reference Data
  - Reproducibility
  - Training materials
- Starting with USA, Europe and Australia, more welcome!
- Manage with community assets/repositories
- Don’t prescribe hardware resources

Community assets
usegalaxy.* - Tools

Shared repo of tool lists
- Minimum toolset
- Extra tools by genre
  - Metagenomics
  - Proteomics
  - Metabolomics
- Curation/maintenance of tools and versions in yaml files
- Automatic upgrades and installation of trusted tools
- Still allow local specialisations
usegalaxy.* - Local Specialisations

Simple
- Specialist tools for an area of research
- Local requests

Complex
- Sub-web site of usegalaxy.eu
- Re-skinned Galaxy
usegalaxy.* - Reference Data/Indices

Tier 1 uses smart caching
- On demand
- Local specialisation

Run by usegalaxy.org staff
- Genomic references
- Tool indices
usegalaxy.* - Look and Feel

usegalaxy.* servers will:

- Run the latest stable Galaxy release
- Present a similar tool list layout
- Access to the same reference data
- Be able to run all of the Galaxy Training Network’s core tutorials
- Share testing/training datasets
Global repository of Tools in containers

- Use CVMFS for smart distribution similar to References
- Singularity containers
- Galaxy just uses appropriate container
- Much easier to manage tool lists and versioning
usegalaxy.* - Future - References

- Move to a community based model
- Improve metadata availability
- Improve reference data provenance
usegalaxy.* - Intergalactic Data Commission

- Begun by Björn and Nate in May
- A git repo for community contributions
- Very early days

IDC - Intergalactic (reference) Data Commission

What the IUC (Intergalactic Utilities Commission) is for Galaxy tools is the IDC for (reference) Data.

This repository is the entry point to contribute to the community maintained CVMFS data repository hosting approximately 6TB of public and open reference datasets.

Currently, we foresee two distinct modes of this repository:

- single data-managers that do not depend on an other data manager, e.g. metablast
- data-managers that depend on other data managers and needs to be executed in a workflow like fashion, e.g. hisat2

The single data-managers are easy to run and all you need is a specification on how to run your data manager in a YAML file. An example is the humann2 data manager.

```yaml
---
data_managers:
- id: toolshed.g2.bx.psu.edu/repos/iuc/data_manager_humann2_database_downloader/data_manager_humann2_database_downloader
  # HUMAnN2 database: Nucleotide database: full changoplasm
  # these params correspond to the Galaxy data manager parameters
  params:
    - 'db|database': 'changeplan'
      - 'db|build': '[[ item ]]'
    items:
      - full
  data_table_reload:
    - humann2_nucleotide_database
```
usegalaxy.* - Proposed Global Architecture

Community Managed
Globally Distributed
Reference
Data/Indices
Tool Containers

usegalaxy.eu
usegalaxy.org
usegalaxy.org.au
usegalaxy.org Compute Architecture (June 2018)
*there will be a quiz on this at the end
usegalaxy.* - Future - BYOC

BYOC now in Australia and Europe
- Melbourne resources for Galaxy Australia
- Czech resources can be used by Galaxy Europe
- Adding others

Goal is to allow user/group to add their compute resources
Dynamic job allocation will know
- Where user is from
- What resources they can use
The ultimate goal!

Sign on in Global Galaxy WebServer!

Job is run where-ever your compute and data is located!
It’s a big effort..

Simon Gladman
Derek Benson
Gareth Price
Anna Syme
Igor Makunin
Nuwan Goonasekera
Christina Hall
Helen van der Pol
Andrew Lonie

Nate Coraor
Enis Afgan
Anton Nekrutenko
James Taylor
Dannon Baker
Martin Čech
Dave Bouvier &
The Galaxy Project Team &
Galaxy Project Community

Björn Grüning
Helena Rasche
Bérénice Batut
Anika Erxleben
Joachim Wolff
Mehmet Tekman
Rolf Backofen
Frederik Coppens
Gildas Le Corguillé
& Other Elixir Members