Galaxy Australia - lessons learnt from running a nationally resourced service
Biosciences: the nature of the Australian research community

30,000 health/biosciences researchers

18,000 health/biosciences HDR students

48,000 health/biosciences PG course work students

(163,000 + 40,000 =) 200,000 health/biosciences UG students

1,000 to 1,500 bioinformatician/computational biologists

=> the IT smarts need to be packaged up and delivered cleanly
Efficiency through Galaxy controlled scheduling
A Brief History in Time

• Genomics Virtual Lab for public and private VLs, with Galaxy as a managed public service
  – Galaxy-Melb, Galaxy-Qld and Galaxy-Tut
    • Tool version alignment / tutorial consequences
    • Duplicate staff roles
    • Tut was re-imaged over and over, no longevity to user accounts

• Rebranded Gal-Qld to Galaxy Australia and moved to org.au
  – Staged migration of Melb to Australia

• Brought online 2 x Pulsar (repurpose Galaxy-Melb allocation and new allocation at NCI Canberra)

• Hybrid training events: 4 run nationally with local facilitators and a lead trainer, covering 11 sites and 1064 participants

• Policy, policy and policy – have a lot to thank Galaxy Europe for forging new ground
Solutions for Data Analysis

- Freeware
  - Galaxy
  - R Studio
  - Command Line / HPC

- Commercial
  - Office
    - Excel
  - Agilent
    - Cartagenia Bench Lab for Molecular Pathology
  - Illumina
    - BaseSpace
  - Qiagen
    - CLC-Bio Suite of Analysis Products
  - ThermoFisher
    - Ion Reporter

Most equivocal solution

Galaxy Australia user numbers (as of Sept 2018 - 2268) as CLC-Bio users is a difference of 750k funds (+750K in kind) vs approx. $11million annual licence fees

This does not include the cost of computers to support CLC-Bio installations
Galaxy Australia – distributed architecture

Melbourne Pulsar
- Head Node m2.xlarge
- Worker Node m1.xlarge
- Worker Node m1.xlarge

Canberra Pulsar
- Head Node m2.xlarge
- Worker Node m1.xlarge
- Worker Node m1.xlarge
- Worker Node m1.xlarge

Brisbane
- Head Node - m2.xlarge
  2 x web handlers
  4 job mules - 1 farm

DB Server - m1.large
  PostgreSQL

GPFS Storage

Workers are minimal GVL image, version controlled, updated, with telegraf. Can be swapped into cluster by draining and re-deploying.
Community Impact Summary

A complete bioinformatics analysis platform with:

- >250k bioinformatics jobs run
- 946 bioinformatics tools
- 213 reference genomes

...and on average 2 new tools or references added weekly on user request

Supporting published research across many areas:

- Cardiac
- Toxicology and Drug Discovery
- Cancer
- Vision
- Agriculture and land management
- Obesity
- Evolution
- Ecology
- Infection and immunity

57 peer reviewed papers
14 Higher degree theses

An active and engaged user community

- 5866 registered users.
- 1537 active users (last 90 days)

User growth 2016 - 2019

Registered users in Australia from:

- 30 Australian Universities
- 22 Medical Research Institutes or Organisations
- 19 Other Research Organisations

Users per domain:

- edu.au
- gov.au
- org.au
- com.au
- edu
- other international

Users represented across:

- 420 organisations
- 49 countries
Connecting Australian Data Repositories with Galaxy Australia

Interactive plots/graphs
Presentation of input data and tools/versions used to user

Non consortium members
Australian Microbiome consortium

Invoke Galaxy Australia API
BYO and upload OTU abundance tables

Calculation of taxonomic diversity
Calculation of relative OTU abundance at each site

NMDS ordination
Plot on Kronos graph
Plot on graph
Plot on graph
Community Impact

Jobs/Month and Total Jobs Run on Galaxy Australia.

- Jobs/Month
- Total Jobs

January 2016 - January 2019
usegalaxy.* - a global platform and support network

Distributed reference data between.* servers
- reduced System Administration per locale
- Australian contribution to global efforts
- users are not restricted to "local" content

Intergalactic Data Commission
- formed in 2018
- regulation, automation and documentation of the CVMFS reference collections
- Australian representation on the IDC

Galaxy Project Executive Steering Committee
- formed in 2019
- Australian representation on the Committee
usegalaxy.* - a global platform and support network

Galaxy Training Network
- Australian content contributions
- Multi-language options
- Simple and comprehensive options
- Many peer reviewed best practices
- Synergised tool set

More usegalaxy.* services
- Growing number of countries and regions are forming their own usegalaxy.* service
- Increasing content development
- Distributing SysAdmin activity
- Thursday’s keynote address: The Development of ASEAN Federated Identity and Login Management, and Galaxy ASEAN Community

Welcome to Galaxy Training!
Collection of tutorials developed and maintained by the worldwide Galaxy community

<table>
<thead>
<tr>
<th>Galaxy for Scientists</th>
<th>Galaxy Tips &amp; Tricks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic</strong></td>
<td><strong>Tutorials</strong></td>
</tr>
<tr>
<td>Introduction to Galaxy Analyses</td>
<td>7</td>
</tr>
<tr>
<td>Assembly</td>
<td>4</td>
</tr>
<tr>
<td>Computational chemistry</td>
<td>4</td>
</tr>
<tr>
<td>Ecology</td>
<td>5</td>
</tr>
<tr>
<td>Epigenetics</td>
<td>5</td>
</tr>
<tr>
<td>Genome Annotation</td>
<td>4</td>
</tr>
<tr>
<td><strong>Topic</strong></td>
<td><strong>Tutorials</strong></td>
</tr>
<tr>
<td>Data Manipulation</td>
<td>7</td>
</tr>
<tr>
<td>User Interface and Features</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Galaxy for Developers and Admins</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic</strong></td>
<td><strong>Tutorials</strong></td>
</tr>
<tr>
<td>Galaxy Server administration</td>
<td>31</td>
</tr>
</tbody>
</table>
Growing our Australian Community – Metabolomics Australia

• Provide vendor agnostic tools and visualisations

• Leverage an increasingly mature global open source community
Growing our Australian Community – Oz Mammal Genomes and Target Capture Panel pipelines

- Build on 2018 experience with BPA Data Portal Australian Microbiome

- User interact at BPA Data Portal (and potentially never leave!) to trigger pipeline on Galaxy Australia

- Analysed data returned to Data Portal and made publicly visible
Growing our Australian Community – authentication and storage links

- Galaxy Project is an open collaborative global project
- Users on public Galaxy services rarely are required to authenticate
- But with authentication comes opportunity for tailored or seamless servicing

- Link Galaxy to AAF
  - Link to institutional HPC destinations

- Link Galaxy to AARNET Cloudstor
  - Data Input
  - Data Export
  - Service provider mediated Input
Galaxy Australia Team Members

Gareth Price          Nick Rhodes
Igor Makunin          Simon Gladman
Thom Cuddihy          Nuwan Goonasekera

Special Thanks:

Sarah Richmond, Ecoscience Research Cloud (ecocloud)
Derek Benson, CSIRO
Anna Syme, Royal Botanical Gardens, AU
Grahame Bowland, QCIF