GALAXY AT THE IFB: FROM COMPUTING TO TRAINING

Victoria Dominguez del Angel & Christophe Caron @ifb_bioinfo
OUTLINE

• EOSC / Challenges in Biology
• IFB overview
• Cloud Computing and Reproducible Tools
• Galaxy at the IFB
BIG DATA IS BECOMING THE NEW "BLACK GOLD"
EUROPEAN OPEN SCIENCE CLOUD

- A virtual environment for Europe's 1.7 million researchers.
SOME IMPORTANT QUESTIONS

- Infrastructure
- Interoperability
- Security
- Business Model
EUROPEAN OPEN SCIENCE CLOUD : DMP

- Security and sustainability of data
EUROPEAN OPEN SCIENCE CLOUD: FAIR

RESEARCH DATA - OPEN BY DEFAULT

Accessible

Findable

Interoperable

Re-usable

HORIZON 2020 GRANTEES ARE REQUIRED

take measures to ensure open access to the data underlying their scientific publications

provide open access to any other research data of their choice

Horizon 2020 grantees are encouraged to also share datasets beyond publication
Institut Français de Bioinformatique

French Infrastructure for Biological Information Sciences

IFB : FEDERATION OF BIOINFORMATIC PLATFORMS
A distributed infrastructure for life-science information

France: 35 platforms
SERVICES OVERVIEW

- Data: provision of curated data collections
- Tools: diffusion of innovative tools for analyzing biological data
- Support
- Training: of the life science users
TRAINING 2016

- 1705 Students (Universities)
- 1497 Professional training courses
- Clouds and VMs in bioinformatics, Genome assembly & annotation and Metagenomics
WORKING GROUPS

- Sustainability
- CoBRAS
- Big Data *
- Galaxy
- GRISBI
Repeat, reproduce, replicate, reuse...

- Repeat
  - same experiment
  - same lab
- Reproduce
  - same experiment
  - different set up
- Replicate
  - different experiment
  - different lab
- Reuse
  - experiment
  - some of same

Drummond C Replicability is not Reproducibility: Nor is it Good Science, online
http://www.slideshare.net/carolegoble/
FAIR PRINCIPLES

- Findable
  - (meta)data assigned a unique persistent identifier
  - registered in a searchable resource
- Accessible
  - (meta)data retrievable using standardised communication protocol
- Interoperable
  - (meta)data use a formal, accessible and broadly applicable language
- Re-usable
  - meta(data) are released with a clear usage license, associated with their provenance and improve community standards
WORKING GROUPS

- Sustainability
- CoBRAS
- Big Data
- Galaxy *
- GRISBI
French Galaxy Working Group

- IFB et al.

- 2013 – 2017 !

- Skilled and nice people

- 2 meetings/year

GCC 2017
3 missions

• **Chair** the Galaxy French Community
  – Events organization: Galaxy Day...

• **Foster** development for communities
  – Build Galaxy based solution with and for communities (Metabolomics, Proteomics...)
  – Intergalactic Utilities Commission
  – ...

• Build/organize **Training** for developers and users
Training activities
Galaxy4Bioinformatics

  - **Training package**
  - Modules: Admin, tool integration, API...
  - Aim: Promote best practices

- A French Galaxy Tour Training
  - 6 sessions since 2 years / 3-4 days
  - Customized sessions for 2 communities
    - Imaging, NGS

GCC 2017
ELIXIR Galaxy Developer Workshop

- Starting point to build an ELIXIR session
- Strasbourg / January 2017
  - Participation (~7 trainers) and co-organization
  - Including Galaxy RockStars : F. Coppens, B. Gruening...

~30 participants
Training for biologists

- Concept: Bring Your Own Data
  - Aim: How to use Galaxy to analyze own data?
- Each participant will be paired with a tutor
  - In order to help him prepare the data and assist him during each analysis step.
- Alternate theoretical and practical sessions

- 2 communities
  - NGS: EBA training
    - 5 sessions / 5 days / Roscoff
  - Metabolomics: Workflow4Experimenters
    - 3 sessions / 5 days / Roscoff + Paris

- IFB: Distributed Galaxy Infrastructure
  - Roscoff, Strasbourg, Toulouse, Paris...

200 participants

75 participants

GCC 2017
Development projects
- **Workflow4Metabolomics**: VRE for the Metabolomics and for large-scale analysis  
  - Gildas & Franck - Thursday / 2:20 pm

- **ProteoRE**: Proteomics Research Environment (annotation of MS-based proteomics data)  
  - Lien - Poster #17

- **Galaxy For Life Science**: Automatic building Galaxy environment  
  - Valentin - Poster #9

- **NG Phylogeny.fr**: innovative phylogenetic services  
  - Damien + Fabien - Thursday / 3:20 pm and Poster #6

- **Mississippi**: Galaxy platform for small RNA analysis  
  - Christophe / no poster 😊
WORKING GROUPS

- Sustainability
- CoBRAS
- Big Data
- Galaxy
- GRISBI*
IFB E-INFRASTRUCTURE

- National infrastructure is hosted at IDRIS-CNRS
- 5000 cores and 1 PB storage
- At the end 2017 the IFB’s infrastructure will be provided: 10,000 cores and 2PB storage
RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD
Catalogue des appliances bioinformatiques dans le cloud, parcourez et filtrez les en utilisant les termes présents dans l'ontologie EDAM, ou en langue naturelle.
DEMO

https://goo.gl/b58ier
THANK YOU FOR YOUR ATTENTION
TALKS AND POSTERS AT IFB

- Workflow4Metabolomics: 2:20 pm
- NG Phylogeny: 3:20 pm / Poster #6
- ProteoRE: Poster #17
- GFLS: Poster #9
- Biosphere: S. Perrin
- Anthony - Deployment of genome databases for insects… Poster #21
- Julien - GalaxyCat: Lightning talk
Questions ?