Building an open, collaborative, online infrastructure for bioinformatics training

Bérénice Batut, Galaxy Training Network, Dave Clements, Björn Grüning

Galaxy Community Conference
June 2017
Why caring about bioinformatics training?
Need for bioinformatic training

Bioinformatics has become too central to biology to be left to specialist bioinformaticians

- Explosion of data to analyze
- Access to computational power
- Thousand of possible tools for specialized analyses
An increasing demand for learning bioinformatics

Graphs of Brazas et al, 2017
Galaxy
a great solution!
Computational knowledge: Not required!

- Web interface for numerous bioinformatics tools
- Scalable
- No issue with computer configuration during training
Quite a lot of events...

Number of registered GTN training events over the months
Building a new **open, collaborative** and **FAIR** model for bioinformatics training
Requirements

- Easy to use
- Support for effective training for
  - Individual users
  - Instructors
- Definition of technological infrastructure
- Limited redundancy
The model
One repository to collect everything

GitHub: galaxyproject/training-material
Topics for different targeted users

- admin
- assembly
- chip-seq
- dev
- introduction
- metagenomics
- proteomics
- sequence-analysis
- trainers
- variant-analysis
Similar structure, content and formats
Separation between content and formatting
Citable data & Credit

Diagram showing a tree structure with directories and files, including:
- Slides
- Images
- Tutorials
- README (Markdown)
- Metadata (YAML)
- Docker
- MotherMiseq-SOP
  - Metadata (YAML)
  - Slides (Markdown)
  - Tutorial (Markdown)
- Tours
- Workflows (YAML)
- Tools (YAML)
- Data Libraries (YAML)

Zenodo

DOI

Input file
Definition of the technical infrastructure

Annotation, configuration & integration

Public instances
Definition of the technical infrastructure
Findable Accessible Interoperable Reusable

model
Findable

TeSS: elixir Training Portal

http://tess.elixir-europe.org/
Accessible

- Online
  
  http://galaxyproject.github.io/training-material/

- Technical support
  - Self-training boxes with Galaxy Docker flavor
  - Annotated public Galaxy instances
Interoperable

- Metadata description in YAML and integrating EDAM ontology
- Content for different targets (workshops and self-training)
- Technical support for different platforms
Reusable & Open

- CC BY 4.0 license
- DOI for the input datasets
- Open development process on GitHub & via Gitter
- Open education movement
Community effort
Numerous contributions

And 2 successful hackathons!
Numerous discussions

Gitter: Galaxy-Training-Network/Lobby
A constantly growing community

Number of new contributors

- GCC 2016
- Online Contribution Fest
- Cambridge Training Hackathon

2015-08 2015-11 2016-02 2016-05 2016-08 2016-11 2017-02 2017-05
0 2 4 6 8 10 12
THANKS to all of them!