Building an open, collaborative, online infrastructure for bioinformatics training

Berenice Batut[1], Galaxy Training Network[2], Dave Clements[3], Bjoern Gruening[1]

[1] *University of Freiburg, Germany*
[2] *https://galaxyproject.org/teach/gtn/*
[3] *Johns Hopkins University, United States of America*
berenice.batut@gmail.com

With the advent of high-throughput platforms, life science data analysis is tightly linked to the use of bioinformatics tools, resources, and high-performance computing. However, the scientists who generate the data often do not have the knowledge required to be fully conversant with such analyses. To involve them in their own data analysis, these scientists must acquire bioinformatics vocabulary and skills through training.

Data analysis training is particularly challenging without a computational background. The Galaxy framework is addressing this problem by offering a web-based, intuitive and accessible user interface to numerous bioinformatics tools.

Recently, the Galaxy Training Network (GTN) set up a new open, collaborative, online model for delivering high-quality bioinformatics training material: http://training.galaxyproject.org.

Each of the current 13 topics provides tutorials with hands-on, slides and interactive tours. Tours are a new way to go through an entire analysis, step by step inside Galaxy in an interactive and explorative way. All material is openly reviewed, and iteratively developed in one central repository by almost 50 contributors. Content is written in Markdown and, similarly to Software/Data Carpentry, the model separates presentation from content. In addition, the technological infrastructure needed to teach is described with a list of needed tools, annotation of public Galaxy instances and Docker images for each topic. The data are also stored in Zenodo and citable via DOI.

All materials are annotated by a rich set of metadata (time and resource estimations) and automatically propagated to ELIXIR's TeSS portal. This approach creates tutorials that are accessible, easy to find and (re)use (FAIR) by individuals and by trainers for workshops.

With this community effort, the GTN offers an open, collaborative, FAIR and up-to-date infrastructure for delivering high-quality bioinformatics training for scientists.