

The coming of the Age of the Platform for bioinformatics ? (And how to keep it open-source).

Cloud BioLinux is a Virtual Machine (VM) that is fully cross-platform and open-source, including a large set of pre-installed bioinformatics tools. Since the start of the project in 2009, our aim was to provide an Infrastructure as a Service (IaaS) solution for bioinformatics on the Cloud. With the integration of the Galaxy web portal and the CloudMan cluster computing framework on the VM, developers have build user-friendly, scalable applications.

Recently we have witnessed the release of commercial Platform as a Service (PaaS) solutions using the Cloud to provide one additional level of abstraction for bioinformatics computing, such as for example Illumina's BaseSpace and DNAnexus. With the PaaS model, bioinformatics developers build applications using the platform, and data storage, computational resources, scalability, are taken care by the commercial provider. With direct sequencer data streaming to the Cloud, intuitive "Web 2.0" application interface for users and app market for developers, such platforms provide a great promise but also a peril for commercial vendor lock-in.

This talk will use as example Illumina's Basespace, and go over the technical details of implementing bioinformatics applications to run natively on this specific platform. This will serve as a leeway for presenting the development pitfalls the can lead to vendor lock-in. On the other hand, given the potential of such a platform for streamlining bioinformatics computing we will discuss a technical design for an open-source PaaS platform where applications can be submitted and be fully functional, while staying transparent, accessible and reproducible by the bioinformatics developer community.