

Practical aspects of implementing the IRIDA system as a solution for One Health bioinformatics analyses

Jeevan Karloss Antony-Samy^{‡§}, Georgios Marselis[‡], Eve Zeyl Fiskebeck[‡], Taran Skjerdal[‡], Camilla Sekse[‡], Karin Lagesen[‡]

[‡] Norwegian Veterinary Institute, Oslo, Norway

[§] Norwegian Institute for Public Health, Oslo, Norway

Corresponding author: Karin Lagesen (karin.lagesen@vetinst.no)

Abstract

Managing sequence data, associated metadata, bioinformatics analyses and results can be challenging. In a One Health context, the challenge is even larger as there are many actors involved, many diverse types of results need to be produced, and the ensuing process data, such as software versions and options have to be tracked for auditing purposes. In addition, results must often be produced rapidly to be actionable, and non-bioinformaticians should be able to perform the the analyses. Therefore, a graphical user interface (preferably web system) with pipelines and visualization tools are needed to do these analyses. The Public Health Agency of Canada has together with other actors developed the web based system IRIDA (<https://www.irida.ca>) which uses Galaxy for analyses. IRIDA comes with a set of pipelines, visualization tools and a project based data management system that allows for fine grained data access control, which satisfies many of the requirements that a One Health bioinformatics platform dictates.

However, as is often the case with a system meant to satisfy high demands, the platform is not trivial to set up and adapt for local use. In our setup, we are using two web servers, two database servers and one file server. The IRIDA web server provides the user interface. The Galaxy web server receives commands from IRIDA, executes the commands and returns results. Each web server has a database that keeps their respective metadata: user information, file locations and results. The actual files are stored on the fileserver. This spoke-and-wheel infrastructure was implemented to ensure minimum disruption of service if a component should go down.

To get the necessary compute resources for this system, we are contracting with the Norwegian Research and Education Cloud (NREC), which offers Infrastructure as a Service (IaaS) services for Norwegian institutions and universities. NREC utilizes template VM images which can be instantiated according to need. The automated

configuration and orchestration of images ensure that we can have dynamic access to resources according to need. This dynamic scaling is accomplished through collaboration with Elixir Norway. They have implemented the Pulse software which can check usage and instantiate and take down virtual machines as needed.

At the Institute, we have spent close to two years on exploring and setting up this system. We have learned that it is important to not underestimate the amount compute resources needed to get a solid setup. However, having enough compute is irrelevant without knowledgeable staff. IRIDA comes with many features, which require considerable prior knowledge to adapt and set up in a local infrastructure. This includes knowledge on webservers, database systems, linux administration and Galaxy systems administration. The complexity dictates that these systems need to be set up and managed by in-house IT trained staff that will be able to tend the system along the way. It is also very important to maintain interactions with the users of the system, to ensure that the setup produces results that are useful to the users. To accomplish this, bioinformaticians are needed to develop pipelines and visualizations that give results that will on their own be easy for users to interpret in a biologically correct manner. Last but not least - such systems require a significant investment from the institution, thus it is important to showcase the benefits that the system will provide.

Keywords

Infrastructure, bioinformatics, genomic epidemiology

Presenting author

Karin Lagesen

Presented at

One Health EJP Annual Scientific Meeting Satellite Workshop 2021 Software Fair

Funding program

One Health EJP has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 773830

Grant title

ORION: One health suRveillance Initiative on harmOnization of data collection and interpretation

Hosting institution

Norwegian Veterinary Institute

Author contributions

JKAS and GM has together with KL set up and tested out the system. KL has worked on project management, development and getting infrastructure resources. EZF and JKAS has evaluated and developed pipelines. TS and CS has tested out and given feedback on the system.

Conflicts of interest

The authors declare no conflict of interest.