A journey of a team of engineers in learning packaging technology

Valentin Marcon1, Laure Quintin2, Durand Patrick2, Olivier Inizan1, Caroline Dussart2, Valentin Loux1, Maria Bernard2, Géraldine Pascal2


PROBLEMATIC

Bioinformatics pipelines installation can take time for engineers and administrators because of the amount of dependencies required.

Facing this problematic, we, a team of four engineers responsible of tools installation where wondering how to facilitate such complex installation for everyone for a pipeline on Galaxy.

STUDY CASE, FROGS: ‘Find Rapidly OTUs with Galaxy Solution’

THE PIPELINE

This 

meta

barcoding

analysis pipeline is designed to give the abundance table and the taxonomic affiliation of operational taxonomic units (OTUs) to be used as a Galaxy workflow, focused on user-friendliness, so it does not require bioinformatics or command lines skills.

WHY FROGS?

20 training
200 people trained
17 citations
~400 visit/month

FROGS.toulouse.inra.fr

DEPENDENCIES

FROGS pipeline is a good case of study because it contains about 20 wrappers (tools in Galaxy) and needs about 20 dependencies to work. Wrappers and source code are host on the Github page of developers and to be installed manually. Dependencies needs to be found in various websites and to be installed manually too.

TECHNOLOGIES

The key issue was that we did not have a good knowledge in packaging technologies. But we didn’t made a huge work of research, we just followed the advices of the Galaxy Community.

The packaging technology provided by Galaxy is Conda, which allow you to fully make a tool ready to use with only one command (as “apt-get install”) and to manage virtual environments. Then, we wanted to publish the wrappers (Galaxy integrated tools) on the Toolshed (the galaxy tools catalog). For that purpose we used Planemo (a tool that help people to create, validate and test their own wrapper) to ensure that the wrappers made by the FROGS developers work correctly where following the best practices of designing wrappers.

Finally, we wanted to figure out whether or not this technology will be easy to learn and if we will be able to package FROGS in a reasonable amount of time.

GOAL

FROGS is accessible for every users, the ones familiar with the command line and the others that are not, thanks to the Galaxy interface. It is very much difficult for the administrators because they have to install all wrappers, sources and dependencies manually. So, our objective was to improve the overall installation process of the FROGS pipeline.

CONCLUSION

Today FROGS pipeline is available as a Conda package through the Biocconda channel, ready to use on the command line.

The Galaxy wrappers are accessible from the Test toolshed in order to install it quickly from your Galaxy administrator interface.

Some issues need to be fixed, and we will publish it soon on the Toolshed. Finally we want to insist on particular advices according to our personal experience. First for the tools packaging: Provide software data that we consider as default dependencies; be careful with Conda channels order; Provide separate Conda packages for complex softwares. Then for the tools development: Provide softwares with tests; Use a logging framework instead of “/dev/null” redirection.

REFERENCES

2. The journey of a team of engineers in learning packaging technology. [Talk on the “Galaxy Community Conference and Bioinformatics Open Source Conference”; web: http://pldstill.com/2015/02/01/galxy-Conference-Community/]

CONCATS

pgdurand@ifremer.fr
laure.quintin@ifremer.fr
olivier.inizan@ifremer.fr
valentin.marcon@inra.fr