The journey of a team of engineers in learning packaging technology.

Laure Quintric¹, Patrick Durand¹, Olivier Inizan², Valentin Marcon², Caroline Dussart¹, Valentin Loux², Maria Bernard³, Géraldine Pascal⁴

¹ IRSI-RIC-Cellule Bioinformatique, IFREMER, 29280 Plouzané, France. Email: Laure.Quintric@ifremer.fr
² INRA, UR1404, MaLAGE, Université Paris-Saclay, Jouy en Josas, France.
³ GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France.
⁴ GenPhySE, Université de Toulouse, INRA, INPT, ENV'T, Castanet Tolosan, France.

Project Website: https://github.com/geraldinepascal/FROGS/
Source Code: https://github.com/bioconda/bioconda-recipes

Abstract:

Imagine that you have a wonderful pipeline that finds, rapidly, OTUs (operational taxonomic units). Imagine that this pipeline meets a growing success in the metagenomic community.

As engineers working on bioinformatics platforms you want to package it and distribute it. Despite the fact that this pipeline is already wrapped for Galaxy (yes, you can Find, Rapidly, OTUs with a Galaxy Solution aka FROGS [Escudie et al., Bioinformatics, 2017]), the point is that this pipeline is built upon on more than 20 dependencies. Again, you are an enthusiast engineer but you are far from being an expert of packaging and deployment.

In this talk we want to relate the journey of packaging and distributing a pipeline that enables a significant amount of dependencies. We are engineers in charge of setting up production quality bioinformatics services for the community of users of several French academic research institutes (IFREMER and INRA). On a daily basis, part of our work consists in installing and configuring softwares and we would like to have these steps as efficient as possible. At the beginning of the journey we had no particular skills in technologies for packaging of softwares and their deployment on a mutualized infrastructure.

We will focus on the learning curve for the technologies and tools employed (CONDA and PLANEMO). We will show how (i) the fact that all team members decided to learn together and (ii) the support and reactivity of the developer community has speed up the learning process. We will also relate the early installation experiences done by people not involved in the packaging and with also no particular skills in dependencies resolution technologies.

Based on this experience, we will give a feedback on some easy-to-implement rules that would greatly simplify tasks to package, test and deploy complex bioinformatics pipelines right into Galaxy.