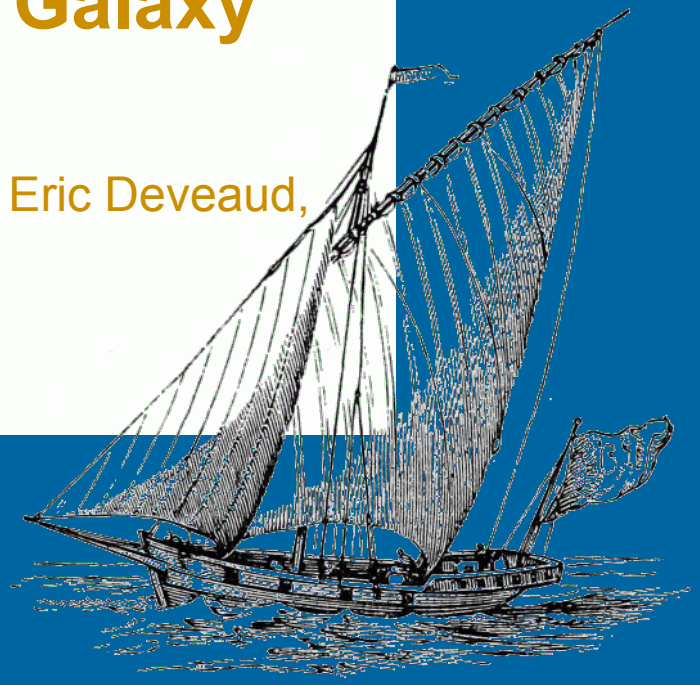
ReGaTE, Registration of Galaxy Tools in Elixir

Olivia Doppelt-Azeroual, Fabien Mareuil, Eric Deveaud, Matus Kalas, and Hervé Menager

07/07/2015



FOR RESEARCH, FOR HEALTH,
FOR OUR FUTURE

UNIVERSITAS BERGENSIS

Institut Pasteur

Plan

INTRODUCTION
Original questions

PART 1
Elixir Registry

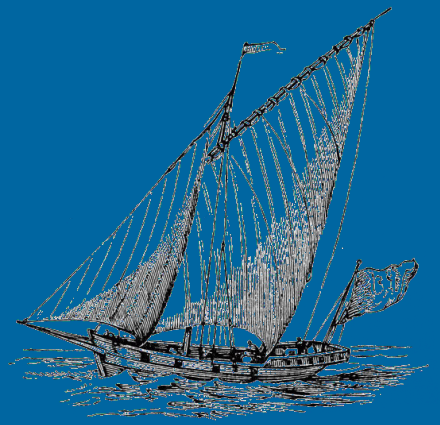
PART 2
ReGaTE

CONCLUSION
& Perspectives

Institut Pasteur

INTRODUCTION

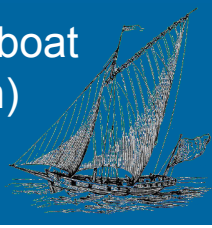
Original questions



Institut Pasteur

1.1 Original Questions

(A regate is a boat race in French)



Data analyst

metagenomic RNAseq

On what kind of project does he work ?

Agent detection proteomics

BAM fastQ VCF

What kind of data does he have/want ?

fasta Krona report

A Galaxy instance A cluster with admins

What ressources does he have access to ?

Not much A virtual machine

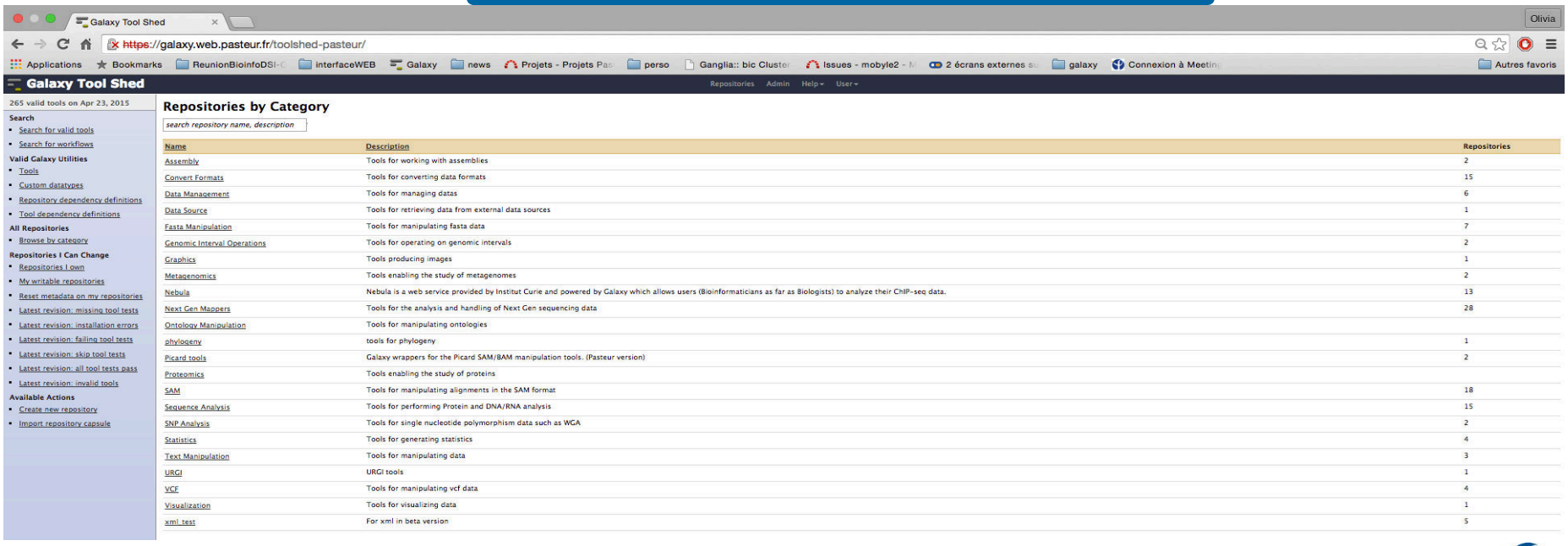
Institut Pasteur

1.2 Start answering, with a galaxyst ?

- if using a Galaxy with admin knowledge ➔ Search in a Toolshed and build analysis workflows but …

WARNING !!

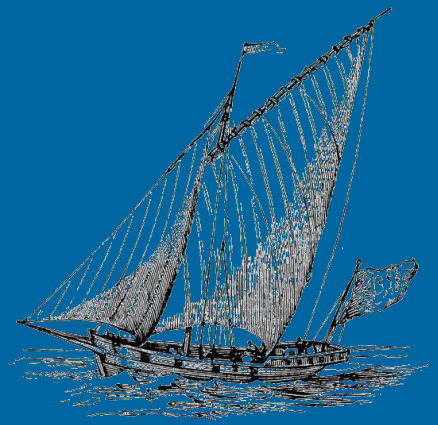
1. HOW ARE TOOLS CLASSIFIED ?
2. NOT ALL TOOLS ARE IN GALAXY



Institut Pasteur

PART 1

Elixir Registry



Institut Pasteur

2.1 ELIXIR Tools and Data Services Registry

Why **A** registry?

A registry addresses the question of resource discovery

➔ Aims in **FINDING** and **UNDERSTANDING** relevant resources by various means

➔ Gives relevant information on that resource

➔ Tells how to access it (web services' url, download pages, …)

Why **THIS** registry?

➔ Uses **EDAM** ontology: a **CONTROLLED** vocabulary to define bioinformatics tools operation, topics, datatypes and data formats.

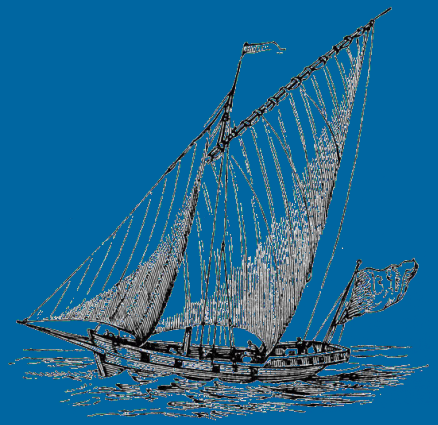
➔ Coupled with workbenchs like Mobyle, Galaxy, … who have become, very useful resources providers,

- Enabling a decentralized registry, maintained by the resource specialist themselves

Institut Pasteur

PART 2

ReGaTE

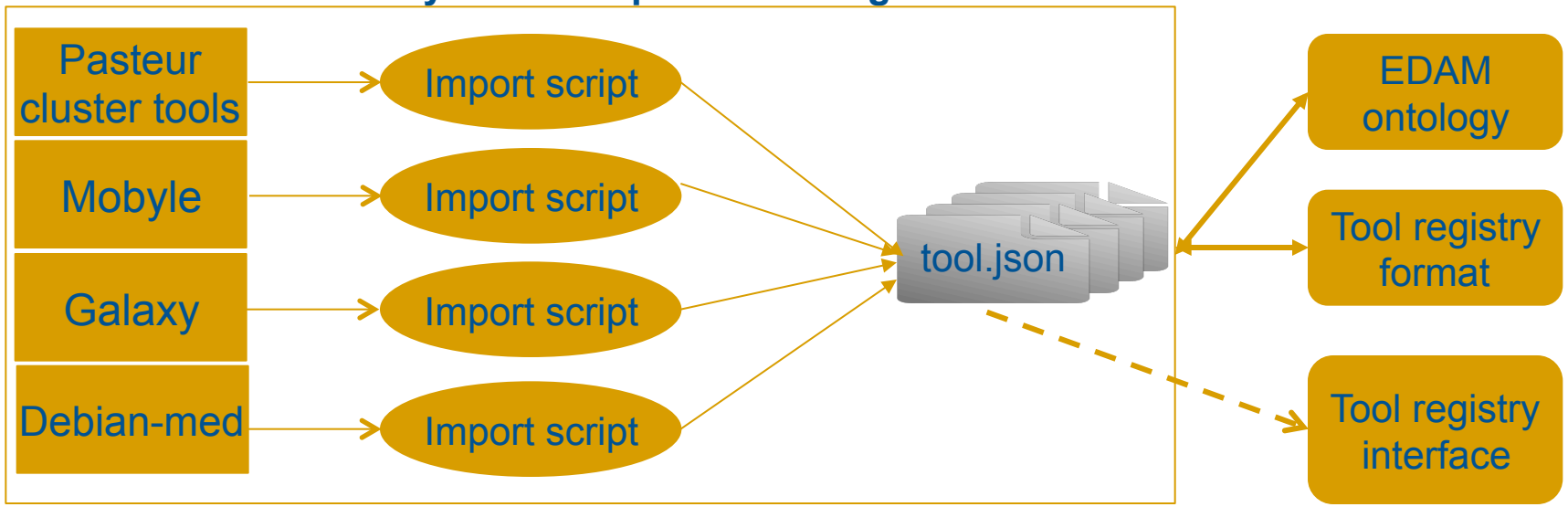


Institut Pasteur

3.1 ReGaTE: History facts

- In June 2014:
  - The toolinfowarehouse subproject was initiated during an EDAM meeting.

➔ AIM: **key resource providers to gather data about bioinformatics tools**



The IDEA:

➔ Use them to start filling “massively” the ELIXIR Registry

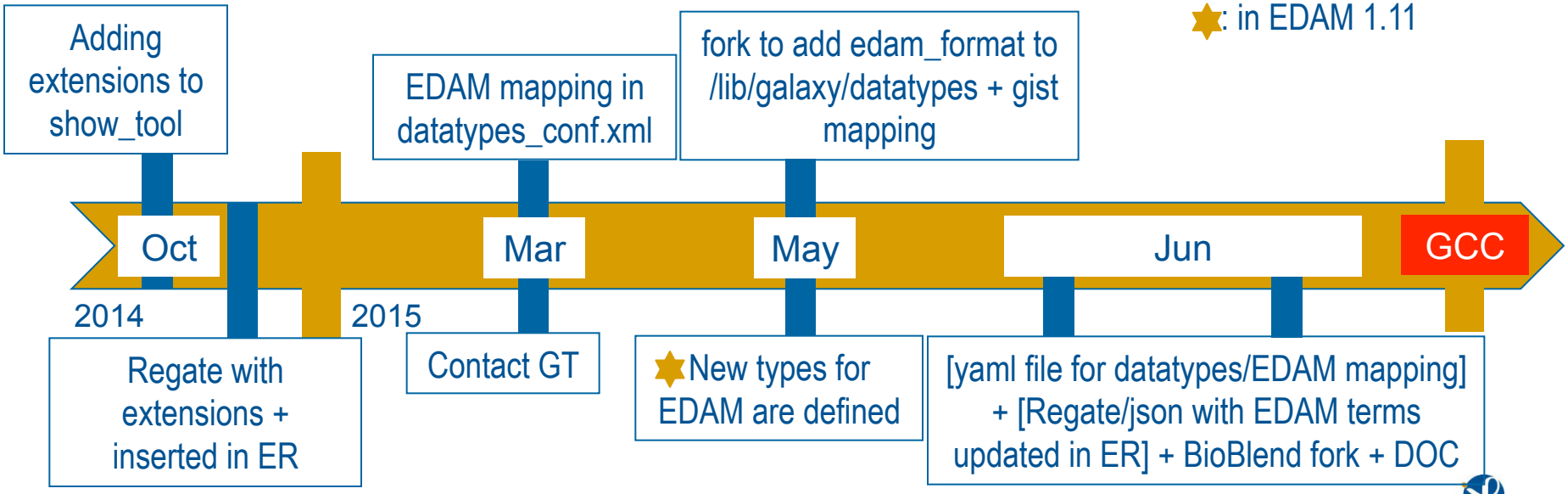
- ReGaTE, stands for Registration of Galaxy Tools in Elixir:
  - Uses Bioblend API to extract information concerning installed tools on any Galaxy; tools.show_tool (io_details=True) to get input/output information
  - Resulting JSON structure files for each installed tool

Institut Pasteur

3.2 ReGaTE: Challenges

- The dictionary returned by the Bioblend show_tool function can be as complex as the Galaxy xml itself
- Galaxy datatypes are a challenge by themselves ☺
- The function field is a key field in Elixir Registry (ER), it gathers:

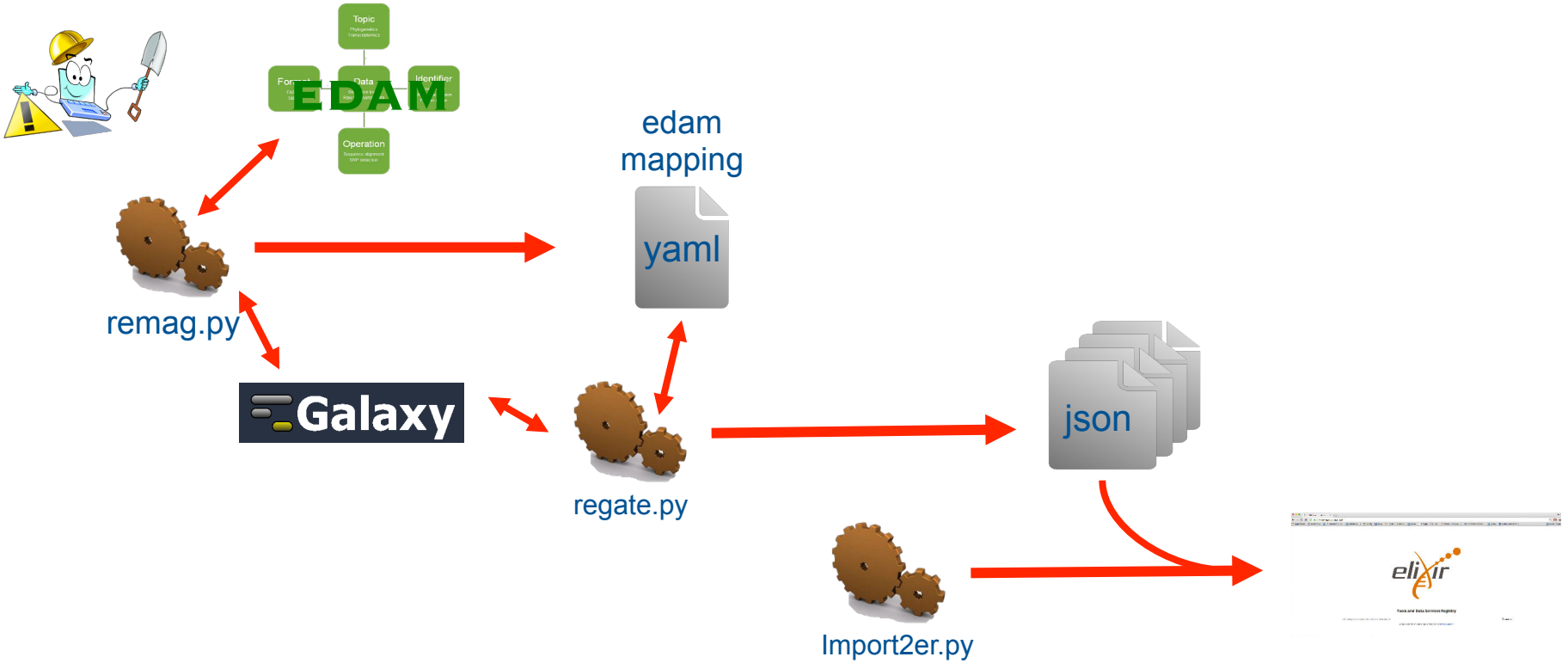
  functionName, functionDescription, functionHandle, input [dataFormat, dataType], output [dataFormat, dataType] ➔ the red fields are EDAM based



Institut Pasteur

3.3 ReGaTE: How it works…

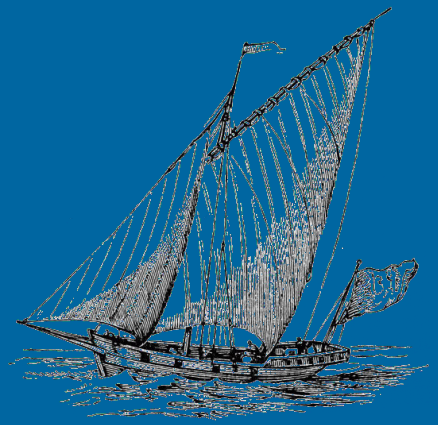
```
##Installation process:
#pip install -e
git+https://github.com/bioinfo-center-pasteur-fr/ReGaTE.git#egg=regate
#pip install -r src/regate/requirements.txt
```



Institut Pasteur

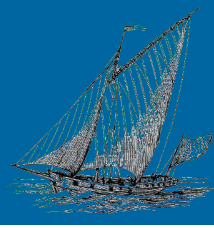
CONCLUSION

& Perspectives



Institut Pasteur

To resume…

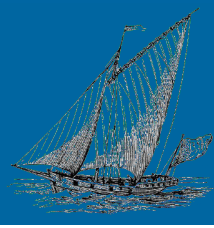


- Some of our jsons are in the registry, key words: “galaxy institut pasteur”
- If interested to display your Galaxy tools in the ER
  - ➔ Use ReGaTE to generate the json files, we tested on several Galaxy servers

Galaxy instance	Nb of jsons	connection	From where
https://galaxyapi.web.pasteur.fr	219	API key	From outside
http://usegalaxy.org	433	API Key	from outside
http://galaxy.sb-roscoff.fr/	377	API Key	ssh on roscoff
http://galaxymetabolomics4api.sb-roscoff.fr/	62	API Key	ssh on roscoff
Your server ?	?	?	?

- You only need a galaxy updated after Oct 2014, an activated API and an API key to enable the use of BIOBLEND
- To insert the JSONs in the Registry, you need to contact the registry group using the mailing list registry@elixir-dk.org.

Conclusion & Perpectives



- The ReGaTE tool is the result of a highly productive collaboration between CBS in Denmark, University of Bergen in Norway, The Institut Pasteur in Paris and the Galaxy Team.

- It is on GiHub: https://github.com/bioinfo-center-pasteur-fr/ReGaTE, with a first version documentation.

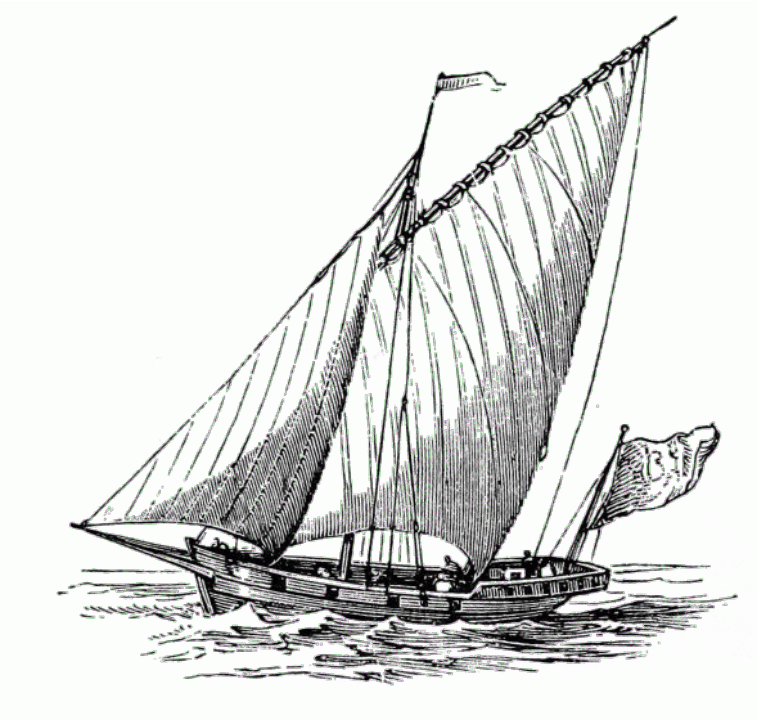
- The addition of EDAM in Galaxy will simplify datatypes management.

- Nice to interact with the Galaxy Team.

- To finish, I'd say that EDAM for galaxy datatypes in only a first step. Indeed, with the French Galaxy workgroup we are planning to work on an "edamization" of the toolshed. (EDAM Operations and Topics to annotate tools)

Galaxy wg Institut Pasteur

THANK YOU !



Institut Pasteur