

# BioMAJ2Galaxy: automatic update of reference data in Galaxy using BioMAJ

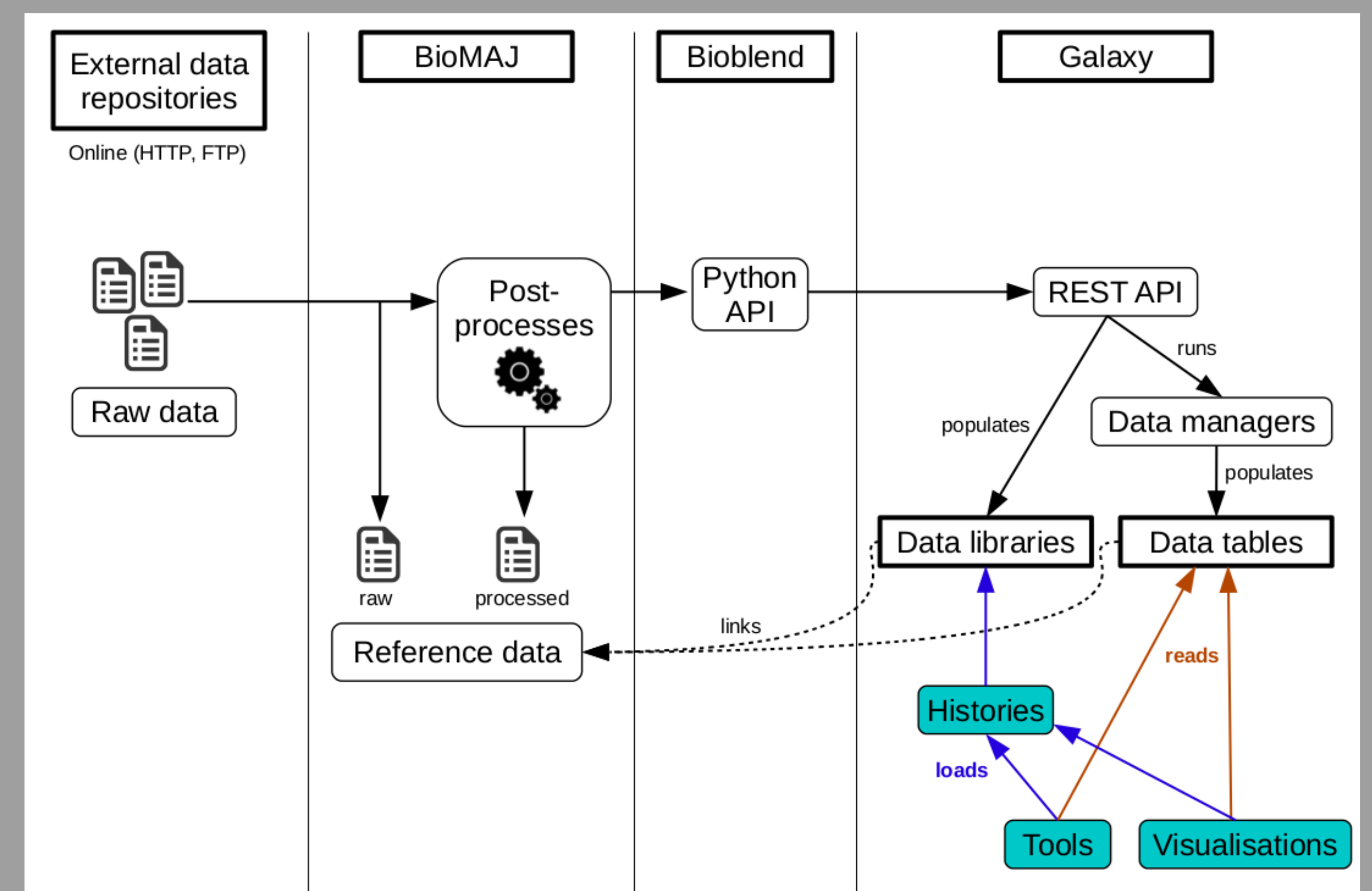
Anthony Bretaudeau<sup>1,2</sup>, Cyril Monjeaud<sup>2</sup>, Yvan Le Bras<sup>2</sup>, Fabrice Legeai<sup>1,3</sup>, Olivier Collin<sup>2</sup>

<https://github.com/genouest/biomaj2galaxy>



## Background: reference data in Galaxy

Many bioinformatics tools use reference data, such as genome assemblies or sequence databanks. Galaxy offers multiple ways to give access to this data through its web interface. However, the process of adding new reference data was customarily manual and time consuming, even more so when this data needed to be indexed in a variety of formats (e.g. Blast, Bowtie, BWA, or 2bit).



The BioMAJ2Galaxy architecture

## BioMAJ

**BioMAJ** is a software that is designed to automate the download and transformation of data from various sources. This data can be accessed directly from the command line, or by using a REST API.

## Implementation

- Uses the Galaxy **API** with **BioBlend** to populate **data libraries** or to invoke **data managers**.
- Several data managers on the GUGGO Tool Shed <http://toolshed.genouest.org>
- Contributions to Galaxy and BioBlend.

Name	Type	Formats	Current Release
Acythosiphon_pisum-bcm	genome/eucaryotic	fasta,blast	2008-06-20
alu	nucleic_protein		2003-11-26
annotqtl	custom		2013-08-15
Anopheles_gambiae-ncbi	genome/eucaryotic	bwa,fasta,bowtie2,blast,2bit	AgamP3
aphidbaseenc	aphidbase/nucleic	fasta,blast	2.1
aphidbaseprot	aphidbase/proteic	fasta,blast	2.1
Apis_mellifera-ncbi	genome/eucaryotic	bwa,fasta,bowtie2,blast,2bit	Arnel_4.5
Arabidopsis_thaliana-ncbi	genome/eucaryotic	bwa,fasta,blast,bowtie2,2bit	TAIR10
Astral		xml	2009-06-03
Bacteria-ncbi	genome/procaryotic	fasta,blast	2013-07-24

The main page of the BioMAJ web interface allowing to configure and monitor databank updates

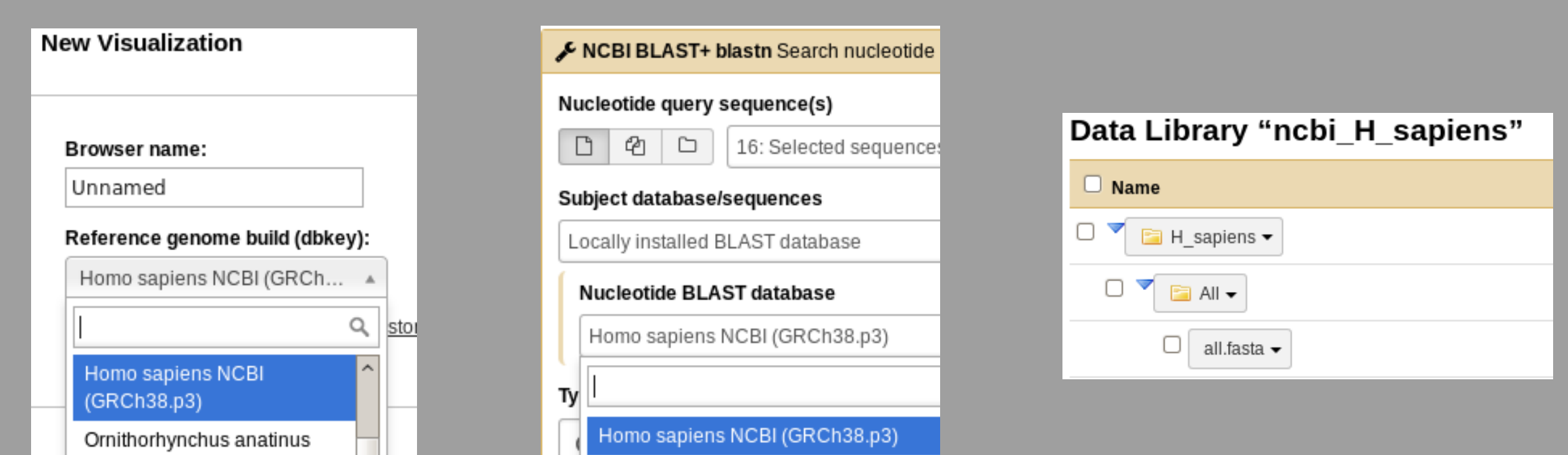
```
B2.db.post.process=GALAXY
GALAXY=galaxy_dm

galaxy_dm.name=galaxy_dm
galaxy_dm.desc=Add files to Galaxy tool data tables
galaxy_dm.type=galaxy
galaxy_dm.exe=add_galaxy_data_manager.py
galaxy_dm.args=-u http://example.org/galaxy/ -k api_key -d "${localrelease}"
-n "Homo sapiens (${remoterelease})" -g ${data.dir}/${dir.version}/${localrelease}
/fasta/all.fasta --bowtie2 ${data.dir}/${dir.version}/${localrelease}/bowtie2/all
--blastn ${data.dir}/${dir.version}/${localrelease}/blast/all

db.remove.process=RM_GALAXY
RM_GALAXY=rm_galaxy_dm

rm_galaxy_dm.name=rm_galaxy_dm
rm_galaxy_dm.desc=Remove from Galaxy tool data tables
rm_galaxy_dm.type=galaxy
rm_galaxy_dm.exe=remove_galaxy_data_manager.py
rm_galaxy_dm.args=-u http://example.org/galaxy/ -k api_key -d "${localrelease}"
-f --blastn --bowtie2
```

Example of a BioMAJ databank configuration using BioMAJ2Galaxy



Reference data are automatically made available in Galaxy

## BioMAJ2Galaxy

To ease the process of giving access to reference data in Galaxy, we developed the BioMAJ2Galaxy module, which enables the gap between BioMAJ and Galaxy to be bridged.

With this module, it is possible to configure BioMAJ to automatically:

- **Download** some reference data
- **Convert** and/or index it in various formats
- **Make available** this data in a Galaxy server using **data libraries** or **data managers**

## Conclusion

These developments allow to integrate the reference data in Galaxy in an **automatic, reliable, and disk-space-saving** way.

The code is freely available:

<https://github.com/genouest/biomaj2galaxy>

GigaScience publication:

doi:10.1186/s13742-015-0063-8

<sup>1</sup> INRA, UMR Institut de Génétique, Environnement et Protection des Plantes (IGEPP), Bioinformatics Platform for Agroecosystems Arthropods (BIPAA), Campus Beaulieu, 35042 Rennes, France

<sup>2</sup> INRIA, IRISA, GenOuest Core Facility, Campus de Beaulieu, 35042 Rennes, France

<sup>3</sup> INRIA, IRISA, GenScale, Campus de Beaulieu, 35042 Rennes, France