

# BioArchLinux: bioinformatics community with Arch Linux

Guoyi Zhang<sup>1</sup>, Yujin Hu<sup>2</sup>, Viktor Drobot<sup>3</sup>, Jens Staal<sup>4</sup>, Yun Yi<sup>5</sup>, Michael G Elliot<sup>6</sup>

1 University of Nottingham, Nottingham, UK. starsareintherose@gmail.com 2 Shenzhen Research Institute of Big Data, Shenzhen, China.

3 Pharm InnTech LLC, Moscow, Russia. 4 Ghent University, Ghent, Belgium. 5 Beijing, China. 6 University of Groningen, Groningen, Netherlands.

## What is BioArchLinux

[BioArchLinux](#) is a **community** for biologists using Arch Linux and Arch Linux-based Linux distributions. It makes a substantial contribution to the Linux ecology for bioinformatics software.

BioArchLinux provides [an excellent repository](#). Uniquely, this repository regularly and automatically checks and updates each package's version based on the latest upstream release, downloading the latest source code from github, sourceforge, cran, or other custom repositories, and compiling new releases for the target architecture. The BioArchLinux repository currently contains nearly 4,000 bioinformatics packages including their dependencies. However, our repository continues to be updated, providing more and more packages, with the goal of matching or exceeding the range of software found in other distribution repositories.

BioArchLinux consists not only of the repository, but also [a wiki site](#), the future goal of which is to provide users with detailed guidance on how to use the software in the repository.

Like Arch Linux itself, the repository is wholly maintained by a voluntary community of enthusiasts and biologists around the world. Users support the community in many aspects, for example by opening an issue, making pull requests and even helping to complete the wiki. In this way, BioArchLinux is a community devoted to serving the users.

## Why choose BioArchLinux?

## Power of community

The PKGBUILD-based packaging method of Arch Linux is much easier than rpm or deb-based Linux distributions. This ease of packaging along with the open community culture are probably the main reasons for the success of AUR (Arch Linux User Repository). There is thus a low threshold for community members to participate in the development of the software available to Arch Linux. This feature is also adopted in our BioArchLinux community. A package can be contributed to the repository with just a single simple shell script and a single YAML script, typically just a few lines long. Thus, after **just one year** of operation we now maintain **over 4,000 packages** with a very small team.

## Keep update and latest

Many Linux distributions can not provide the latest software, especially for software that is frequently updated. In such cases, the latest features can only be obtained by compiling from source code. This comes with the obvious drawback of having to keep track of the package's version and its dependencies (which must also be installed manually).

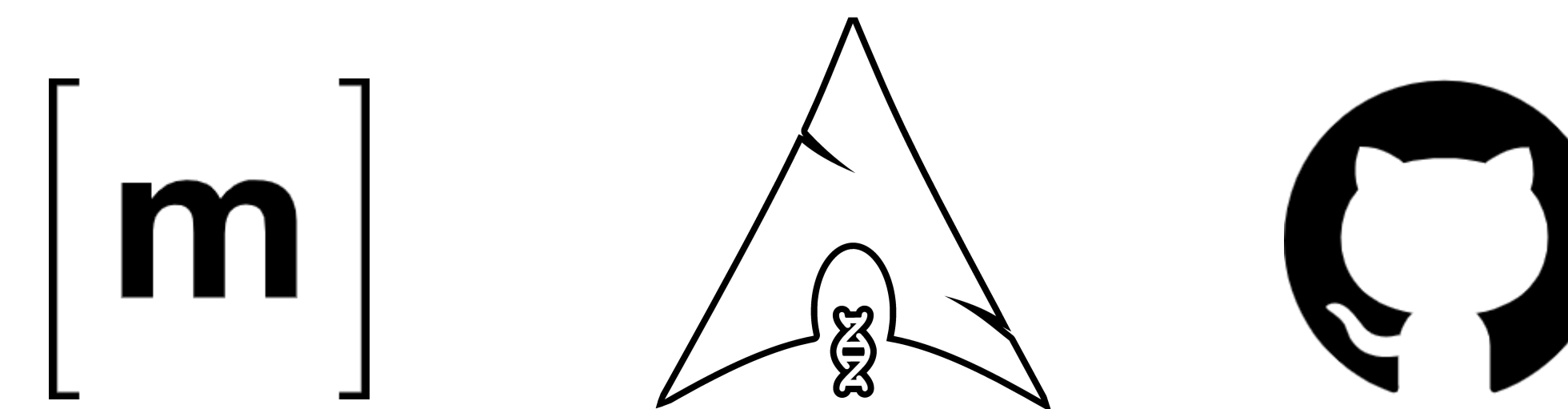
BioArchLinux can help you to skip this step. The repository uses `nvchecker` to **find the latest version** and **build packages as soon as possible** after they have been updated by the author. Therefore, users can just focus on the data analysis using the latest available version.

```
# vim /etc/pacman.conf
-----
[bioarchlinux]
Server = https://repo.bioarchlinux.org/$arch
-----
# pacman-key --recv-keys B1F96021DB62254D
# pacman-key --finger B1F96021DB62254D
# pacman-key --lsign-key B1F96021DB62254D
# pacman -Sy
# pacman -S foo
```

## Usage of BioArchLinux

```
$ pacman -Ss 10.1093/molbev/mst010
bioarchlinux/mafft 7.505-1
    Multiple alignment program for amino acid or
    nucleotide sequences.
    https://doi.org/10.1093/molbev/mst010
```

*Search packages by DOI*



## User friendly

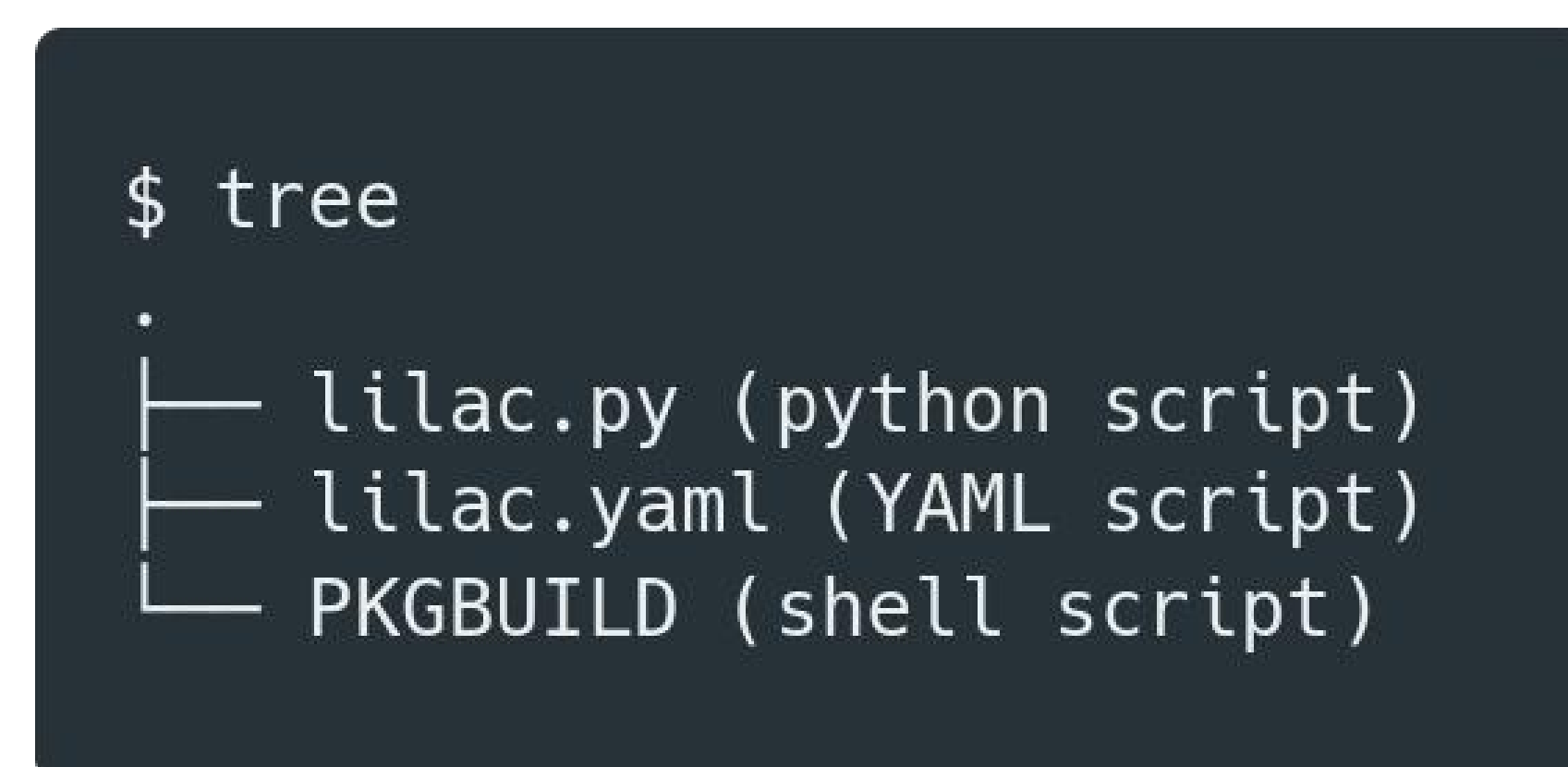
**Installing software from BioArchLinux is easy.** Only several simple commands to install your desired package. Circular dependencies will not be encountered.

The repository provides, **where possible, a DOI in each package description.** This makes it easy for users to find out more about each package’s use and methodology, and to quickly identify appropriate citations when preparing publications.

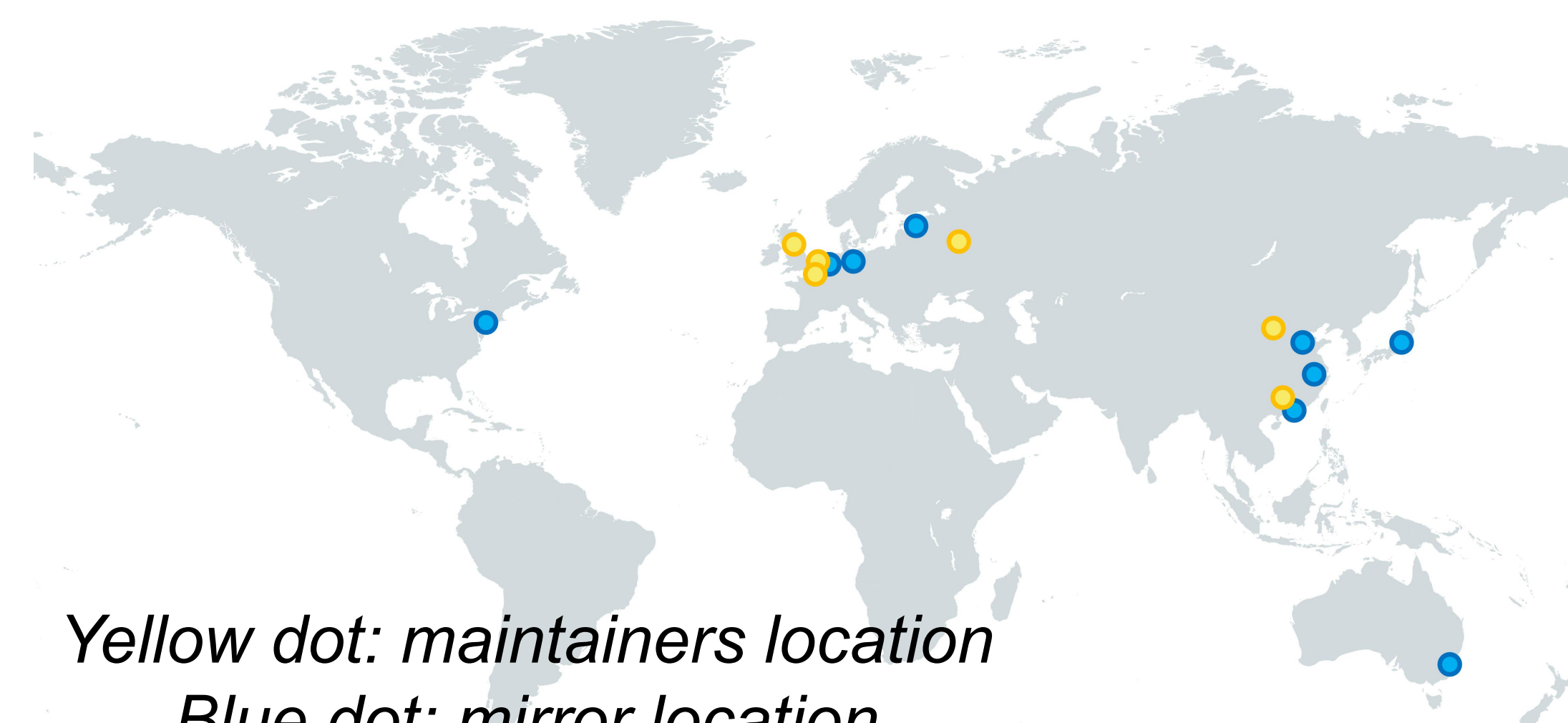
People from countries with internet blocks don't need to suffer slow internet speed and legal risks to access crucial software.

**Local mirrors of our repository would provide great support to their studies.**

**Matrix chat channel** and **GitHub discussions** are open for the users to communicate with the maintainers to solve any possible problems. The wiki site should also give the answer to any common problems.

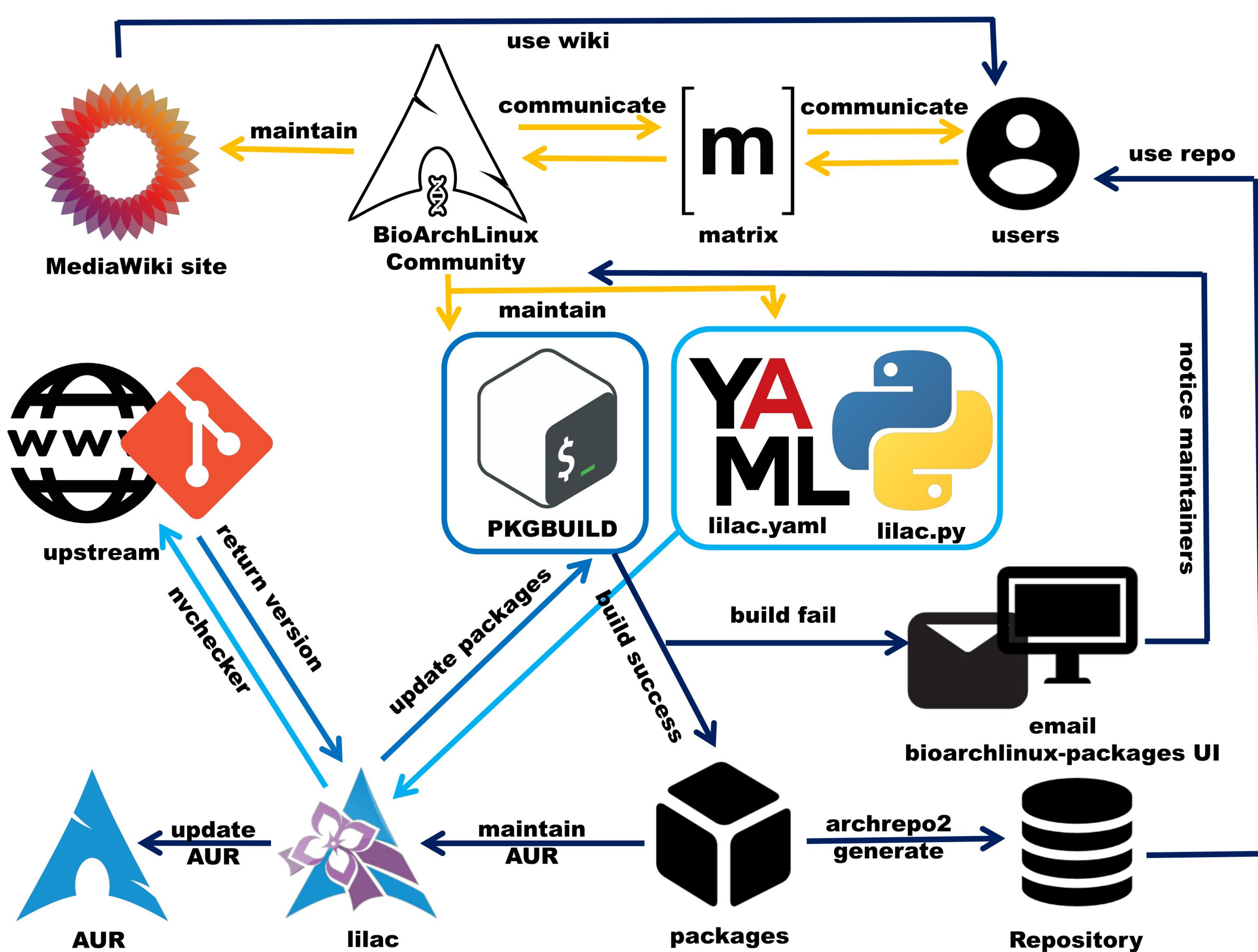


### Package folder structure



Yellow dot: maintainers location  
Blue dot: mirror location

## How BioArchLinux run?



The BioArchLinux repository is maintained by several open source software packages. The main tool is a python application named lilac.

The most basic step is the writing of scripts following the standard of Arch Linux and lilac.yaml. We write a PKGBUILD shell script and a YAML file (and optionally, a python script) and put these in a folder in the git repository.

The nvchecker (new version checker) reads the lilac.yaml to obtain information on the upstream website where the latest version can be checked. If nvchecker can't find the package version from the upstream website, it sends an email to administrators to report the problem.

The information from `nvchecker` is sent to `lilac`, which judges if the package needs to be upgraded or not. If the packages need an upgrade, `lilac` will send the packages to `devtools`, the Arch Linux packaging tool. `Devtools` give the package a clean environment with just the dependencies list in the `PKGBUILD` shell script to allow building. This allows missing dependencies during usage to be avoided.

If the package fails to build, a warning email is automatically sent to the package maintainer. If the packages are built successfully, archrepo2 will put the Arch Linux packages into a specific path and generate a new database file to form a brand new package repository. The entire building process is recorded as log files that can be read using a rust application, bioarchlinux-packages, and shown on [the log website](#).

Our wiki website is built based on MediaWiki. All can freely contribute to this site on the usage of bioinformatics software and bioinformatics concepts and terms.