

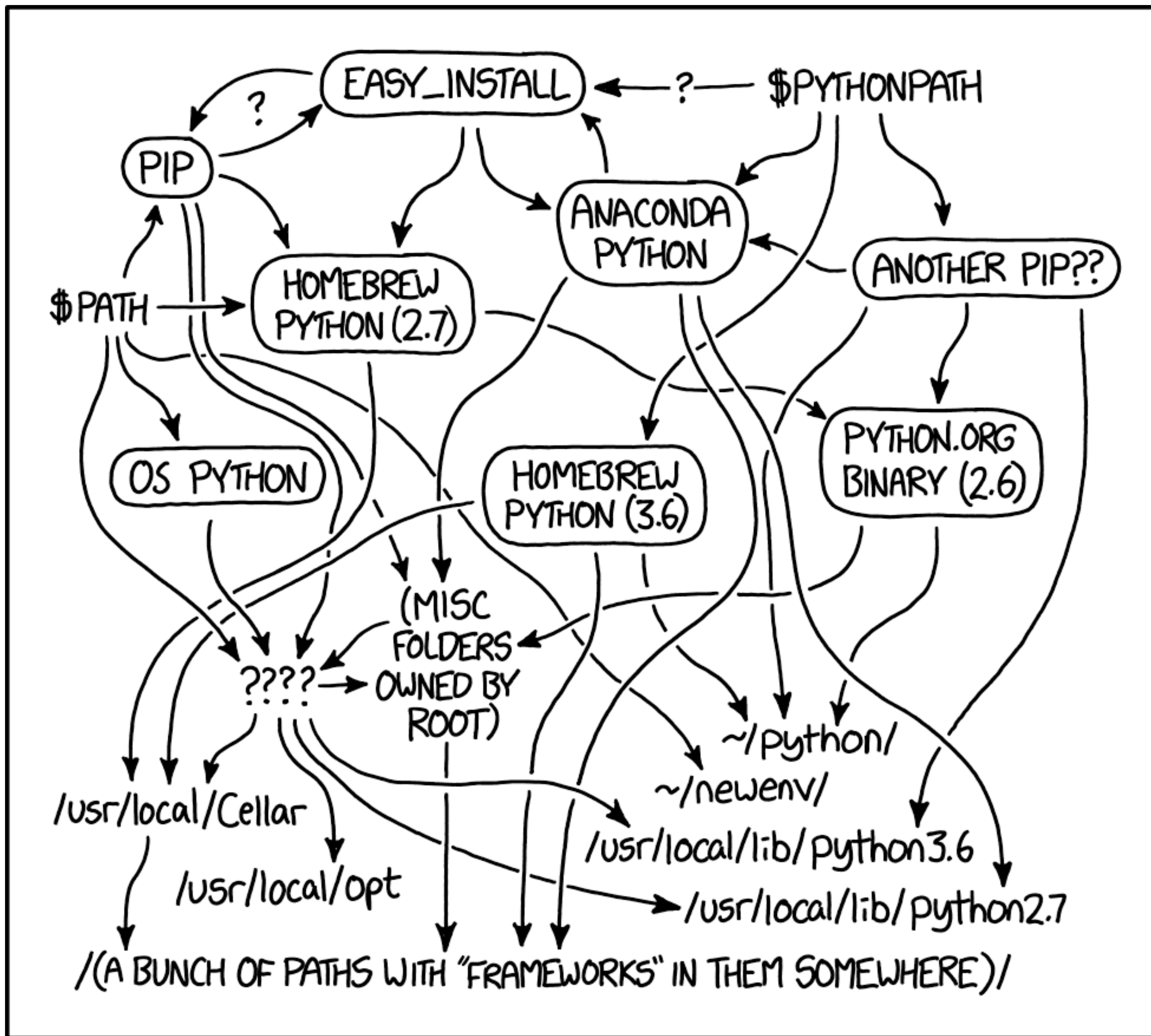
# Packaging for Python and Beyond



Chris Burr

PyHEP 2019 ◦ 16th October 2019





MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.





- PyPI is the main repository of Python packages
  - Almost 200,000 packages
  - Almost 1.5 million releases
- Anyone can claim a name and upload without any review



- ▶ pip is the standard way to install packages
- ▶ Should be as simple as:

```
pip install PACKAGE_NAME
```



- ▶ pip is the standard way to install packages
- ▶ Should be as simple as:

```
pip install PACKAGE_NAME
```



- ▶ You should not:

```
sudo pip install PACKAGE_NAME
```

```
pip install --user PACKAGE_NAME
```





- Standard library module to create environments in isolated folders

```
python -m venv /path/to/new/virtual/environment
```

```
source /path/to/new/virtual/environment/bin/activate
```

- No special privileges required
- Easy to manage many independent environments













## Why virtual environments aren't enough

- Dependent on having a suitable existing Python installation
- Dependent on the upstream maintainers having good packaging
- Incomplete: uses parts of the host system, especially its compilers



# Why virtual environments aren't enough

- Dependent on having a suitable existing Python installation
- Dependent on the upstream maintainers having good packaging
- Incomplete: uses parts of the host system, especially its compilers
- Limited to PyPI packages (though this is evolving: QT and CMake are now available)



- Language agnostic package manager (Python, C++, R, Julia, Rust, Go, Java, Ruby, Fortran, ...)
- Multi platform (Linux, macOS, Windows)
- Multi architecture (i386, x86\_64, aarch64, ppc64le)
- Usage is similar to venv

```
conda create --name my-environment python=3.7 ipython numpy jupyterlab  
conda activate my-environment
```

- Anaconda, Inc. provides around 2,000 packages





- ▶ Language
- ▶ Multi plat
- ▶ Multi arch
- ▶ Usage is s

```
conda create  
conda activate
```



- ▶ Anaconda, Inc. provides around 2,000 packages





- Community maintained collection of conda packages
  - Over 1,600 members
- Over 8,000 packages available and rapidly growing



conda install <package>





- ▶ Community maintained collection of conda packages
  - ▶ Over 1,600 members
- ▶ Over 8,000 packages available and rapidly growing
- ▶ Now includes:





- Language agnostic!
  - Provides it's own compiler toolchain with full C++17 support (Clang 9/GCC 7.3)
  - “Hard” packages are already present and maintained (build systems, Boost, X11)
- It's not just analysis tools
  - Install bash, htop, vim, singularity
- Effectively no dependencies on the host system
  - MacOS 10.9+/CentOS 6+/Windows 64bit
- Open community
  - Missing package: You can add it
  - Broken/hard to install package: You can fix it (and people are generally happy to help)



- Reliably install ROOT in under 5 minutes on any machine
  - Linux, macOS, and Windows Subsystem for LINUX (cling doesn't support native 64-bit Windows)
- Complete installation with C++17, graphics, OpenGL
- Seamlessly integrates with the rest of conda-forge
  - No builtin dependencies (excluding cling for now)
  - No PYTHONPATH/LD\_LIBRARY\_PATH mess
- Downloaded over 50,000 times since it was released 9 months ago
- Currently working with the ROOT team to integrate with their nightlies
  - Plan to release the binaries to a dedicated conda channel
  - If you've be interested in using these, please let me know!





The background is a dense, intricate pattern of blue and yellow lines and shapes. The blue lines form a complex network of straight and curved paths, some resembling a map or a circuit board. The yellow areas are filled with small, dark blue dots and larger, irregular shapes. The overall effect is a highly detailed, abstract design. A white rectangular box is centered horizontally and vertically, containing the text "How does conda-forge work?".

How does conda-forge work?



# Adding a new package to conda-forge

61 lines (53 sloc) | 1.21 KB

Raw Blame History

```
1 {% set name = "zfit" %}
2 {% set version = "0.3.6" %}
3
4 package:
5   name: "{{ name|lower }}"
6   version: "{{ version }}"
7
8 source:
9   url: https://pypi.io/packages/source/{{ name[0] }}/{{ name }}/{{ name }}-{{ version }}.tar.gz
10  sha256: 26e76eb100c95ed52241f3b552d7dd16f59091a83f5e01b263f6fa9f12b30cfe
11
12 build:
13   number: 0
14   script: "{{ PYTHON }}" -m pip install . -vv "
15   noarch: python
16
17 requirements:
18   host:
19     - pip
20     - python >=3.6
21     - setuptools_scm
22     - setuptools_scm_git_archive
23   run:
24     - python >=3.6
25     - tensorflow-base >=1.14.0
26     - tensorflow-probability >=0.6.0
27     - scipy >=1.2
28     - uproot
29     - pandas
30     - numpy
31     - iminuit
32     - typing
33     - colorlog
34     - texttable
35     # Workaround for https://github.com/conda-forge/tensorflow-probability-feedstock/pull/11
36     - decorator
37     - cloudpickle >=0.6.1
38
```

- Create a pull request against <https://github.com/conda-forge/staged-recipes>
- Can be mostly automated using `conda skeleton pypi zfit`

```
39 test:
40   imports:
41     - zfit
42     - zfit.core
43     - zfit.minimizers
44     - zfit.models
45     - zfit.util
46     - zfit.ztf
47
48 about:
49   home: https://github.com/zfit/zfit
50   license: BSD-3-Clause
51   license_family: BSD
52   license_file: LICENSE
53   summary: Physics extension to zfit
54   doc_url: https://zfit.readthedocs.io/
55   dev_url: https://github.com/zfit/zfit
56
57 extra:
58   recipe-maintainers:
59     - chrisburr
60     - mayou36
```



conda-forge / phasespace-feedstock

Sponsor Unwatch 6 Star 0 Fork 1

Code Issues 0 Pull requests 0 Actions Projects 0 Security Insights

## phasespace v1.0.4 #1

Merged chrisburr merged 2 commits into conda-forge:master from regro-cf-autotick-bot:1.0.4 2 days ago

Conversation 1 Commits 2 Checks 2 Files changed 2 +3 -2

regro-cf-autotick... commented 2 days ago

It is very likely that the current package version for this feedstock is out of date.  
Notes for merging this PR:

1. Feel free to push to the bot's branch to update this PR if needed.
2. The bot will almost always only open one PR per version.

Checklist before merging this PR:

- Dependencies have been updated if changed
- Tests have passed
- Updated license if changed and `license_file` is packaged

Note that the bot will stop issuing PRs if more than 3 Version bump PRs generated by the bot are open. If you don't want to package a particular version please close the PR

If this PR was opened in error or needs to be updated please add the `bot-rerun` label to this PR. The bot will close this PR and schedule another one.

Reviewers

- chrisburr
- mayou36

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

```

@@ -1,13 +1,13 @@
1 1  {% set name = "phasespace" %}
2  - {% set version = "1.0.3" %}
3  + {% set version = "1.0.4" %}
4  4  package:
5  5  name: "{{ name|lower }}"
6  6  version: "{{ version }}"
7  7
8  8  source:
9  9  url: https://github.com/zfit/phasespace/archive/{{ version }}.tar.gz
10 - sha256: ad1b322c1c47378ec6687c1ec30d5b92101c883e96bbebea345526b21596516e
10 + sha256: 18e709a27111f96276aaa1f0df073e4cefc5e764ed9551de24b345aa3cc88790
11 11
12 12 build:
13 13 number: 0

```

- Bots monitor for new releases
- Even works with non-standard URLs
- Maintainers normally just have to click merge



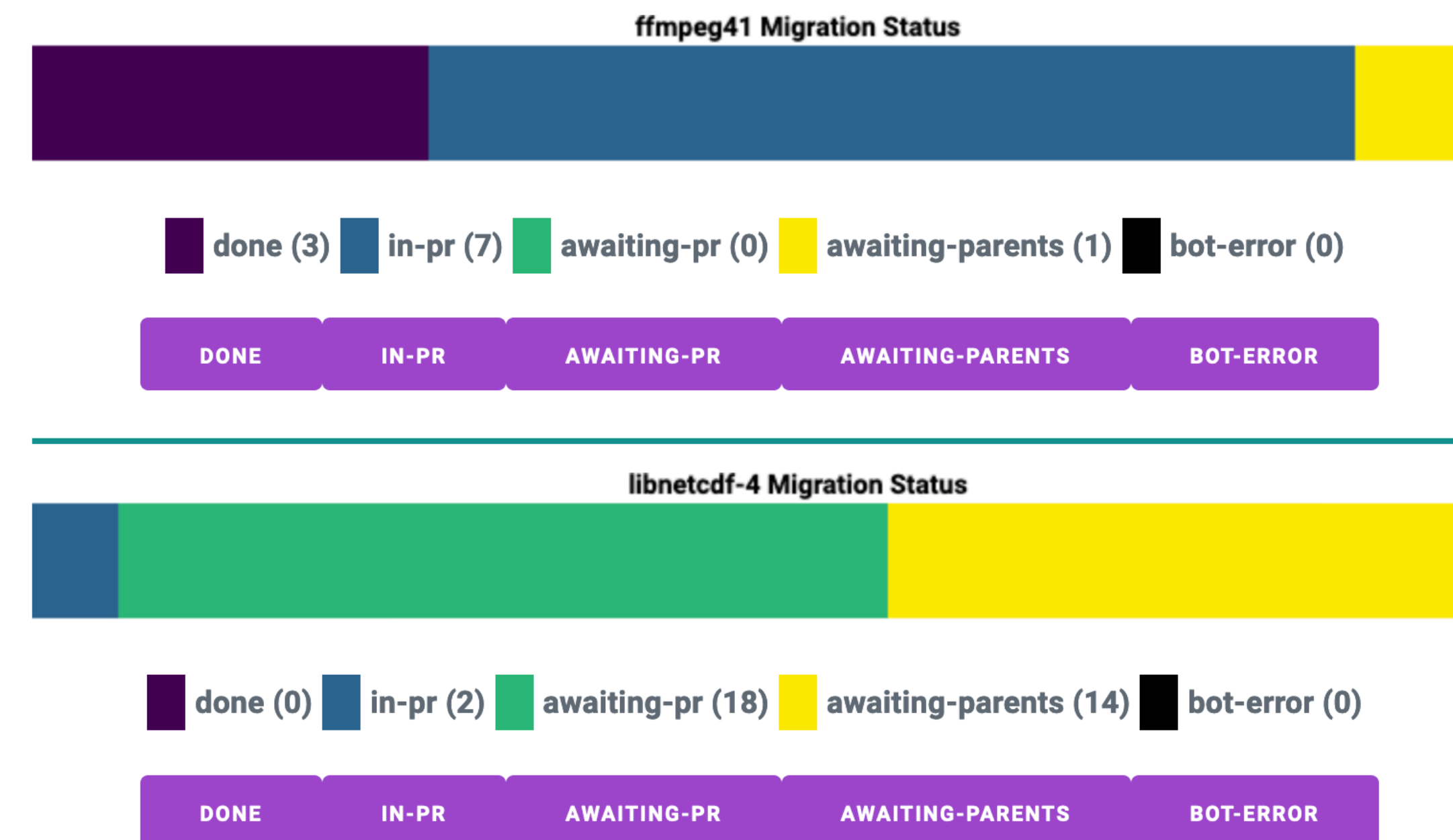
- Conda only supports installing binaries\*
- Relies on the solver knowing about API/ABI compatibility
- Packages with shared libraries should specify what their ABI stability is
- Doesn't necessarily restrict what you can do
  - Variants can be used to provide different versions of a packages
  - BLAS can be provided by netlib, mkl, blis and openblas
  - Several MPI variants
  - TensorFlow has CPU and (several) GPU variants

*\*Some organisations mirror the conda-forge build infrastructure for their own internal use*



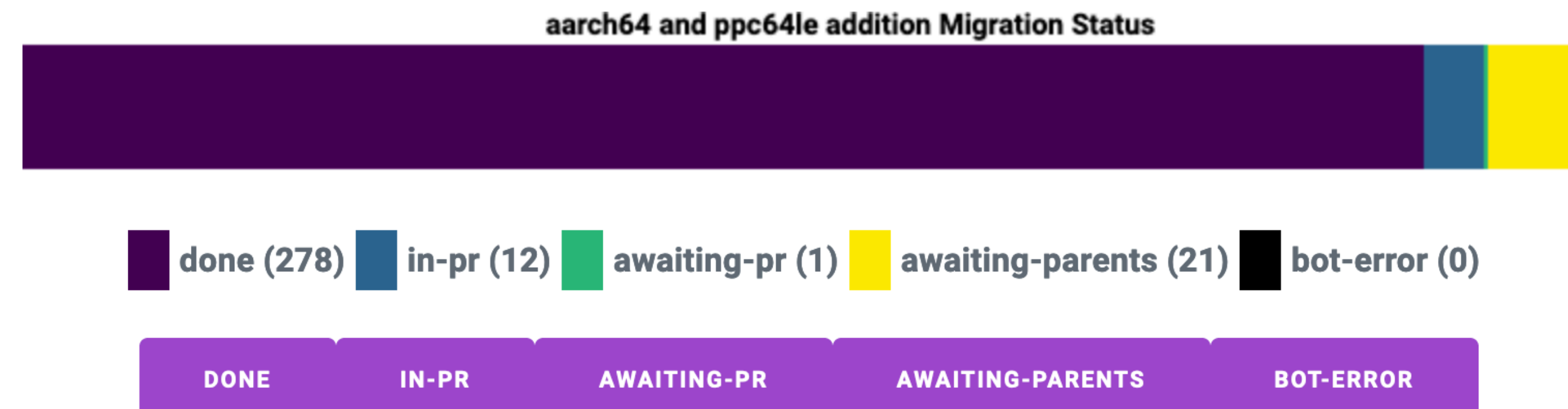
- What about when ABIs change? More 🤖!
- A line is added to a git repository
- Pull requests appear that rebuild packages in the correct order

Current Migrations:





- Migration is currently ongoing for ppc64le and aarch64 support
  - ROOT is included as a target



- Support for compiling CUDA with nvcc is rapidly maturing
  - Adds three additional additional targets (different driver versions)
  - Close to being fully supported by the conda-forge tooling
  - GPU variants of packages are already being added





- One repository per package (“feedstock”)
- All packages are built using well known CI providers
- A year ago this was Travis CI + Circle CI + Appveyor
- Now mostly Azure Pipelines
- Native builds for alternative architectures:
  - ppc64le with Travis CI
  - aarch64 with Drone CI
- All managed by an external package: conda-smithy
  - Used to regenerate CI configuration for each update



- ▶ Long term reproducibly is important for science
- ▶ Easy to export exact builds of everything that was installed

```
conda list --explicit --md5 > environment.txt
```

```
# This file may be used to create an environment using:  
# $ conda create --name <env> --file <this file>  
# platform: osx-64  
@EXPLICIT  
https://repo.anaconda.com/pkgs/main/osx-64/blas-1.0-openblas.conda#4ff605b9a6c88bbfd4428e6f9703d9ff  
https://conda.anaconda.org/conda-forge/osx-64/bzip2-1.0.8-h01d97ff_1.tar.bz2#8397e58ec046d7961aca25e741fa175a  
https://conda.anaconda.org/conda-forge/osx-64/ca-certificates-2019.9.11-hecc5488_0.tar.bz2#c2dd037b660ef2c2b7c5c5535ec1a09b  
https://conda.anaconda.org/conda-forge/noarch/conda-forge-pinning-2019.10.01-0.tar.bz2#49710c701b5be98e25bce22d150a0471  
https://conda.anaconda.org/conda-forge/osx-64/fribidi-1.0.5-h01d97ff_1002.tar.bz2#d5db22ff1a37c085f1f50999facb755c  
https://conda.anaconda.org/conda-forge/osx-64/giflib-5.1.7-h01d97ff_1.tar.bz2#796a4523450c9ab1b495c5c47b10c859  
https://conda.anaconda.org/conda-forge/osx-64/jpeg-9c-h1de35cc_1001.tar.bz2#bcc9abfebf1cc26568e1ec4502834512  
https://conda.anaconda.org/conda-forge/osx-64/libcxxabi-9.0.0-0.tar.bz2#39d20d41d2ecf4d0e8eb626c855b0bef  
https://conda.anaconda.org/conda-forge/osx-64/libgfortran-3.0.1-0.tar.bz2#d69b2c1cc8250395702768acf3a912e3  
https://conda.anaconda.org/conda-forge/osx-64/libiconv-1.15-h01d97ff_1005.tar.bz2#ec331659b7d7ec6565a54bf1ca65ac82  
https://conda.anaconda.org/conda-forge/osx-64/libsodium-1.0.17-h01d97ff_0.tar.bz2#3dd1aeaa242c631c14e4b3b144669fa7  
https://conda.anaconda.org/conda-forge/osx-64/libuv-1.32.0-h01d97ff_0.tar.bz2#fcbcb890e331687f5b118371d3073464  
https://conda.anaconda.org/conda-forge/osx-64/llvm-openmp-9.0.0-h40edb58_0.tar.bz2#7dc368e4f1c6bf92a6dd5e50e25942d9  
https://conda.anaconda.org/conda-forge/osx-64/lzo-2.10-h1de35cc_1000.tar.bz2#9cc86637a9a6ecc9005376afc5b296e0  
https://conda.anaconda.org/conda-forge/osx-64/pandoc-2.7.3-0.tar.bz2#9d7c8563365d39806e683227bc6cfa25  
https://conda.anaconda.org/conda-forge/osx-64/perl-5.26.2-haec8ef5_1006.tar.bz2#7c0ced1da3e20d98b1045f3e945d3d1c  
https://conda.anaconda.org/conda-forge/osx-64/pixman-0.38.0-h01d97ff_1003.tar.bz2#db285954d969a73e3adc622d0db30605
```

```
conda create --name my-new-env --file environment.txt
```



# The problem with long term reproducibility

- Bad metadata in old packages can “poison” the solver
  - Mechanism exists for patching this but it’s fiddly
- Current solution is to move packages to a “broken” channel
  - Channel list must have “conda-forge/label/broken” appended
  - URL also changes (for now)
- Lots of benefits to having a docker container “just in case”
  - Currently setting up a mirroring proxy for conda within LHCb
  - Also deploying to CVMFS
  - Ask me if you’re interested



➤ Creates a relocatable self extracting archive of a conda environment

```
jcrist computer_one $ source activate example
(example) jcrist computer_one $ # Package the current environment
(example) jcrist computer_one $ conda-pack
Collecting packages...
Packing environment at '/Users/jcrist/anaconda/envs/example' to 'example
[#####] | 100% Completed | 5.2s
(example) jcrist computer_one $ ls
example.tar.gz
(example) jcrist computer_one $ # The environment is packaged as a tar.gz
(example) jcrist computer_one $ # Get the file size
(example) jcrist computer_one $ du -h example.tar.gz
55M    example.tar.gz
```

```
jcrist computer_two $ # Activate the environment
jcrist computer_two $ source myenv/bin/activate
(myenv) jcrist computer_two $ # Use applications in the environment
(myenv) jcrist computer_two $ which ipython
/Users/jcrist/computer_two/myenv/bin/ipython
(myenv) jcrist computer_two $ ipython
Python 3.6.5 |Anaconda, Inc.| (default, Apr 26 2018, 08:42:37)
Type 'copyright', 'credits' or 'license' for more information
IPython 6.4.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: import numpy
```

<https://conda.github.io/conda-pack/>



The background is a dense, intricate pattern of blue and yellow lines and shapes. The pattern consists of a complex network of thin, dark blue lines that form various geometric and organic shapes, including spirals, loops, and star-like patterns. The spaces between these lines are filled with a bright yellow color. The overall effect is that of a highly detailed, abstract map or a complex network diagram. A central white rectangular box with a thin black border contains the text "For library maintainers...".

For library maintainers...



- Automate deployment
  - Keep it simple if possible
- Make it easy to run tests against a pre-existing installation
- If it's compiled: provide as many pre-build wheels as you can
  - pip install should work for Python packages
  - But also include source distributions!



- Use `setuptools_scm`
  - Removes all duplication of version numbers

```
setup(  
    name='cirun',  
    use_scm_version=True,  
    setup_requires=['setuptools_scm'],
```

- Uses version control to compute version numbers

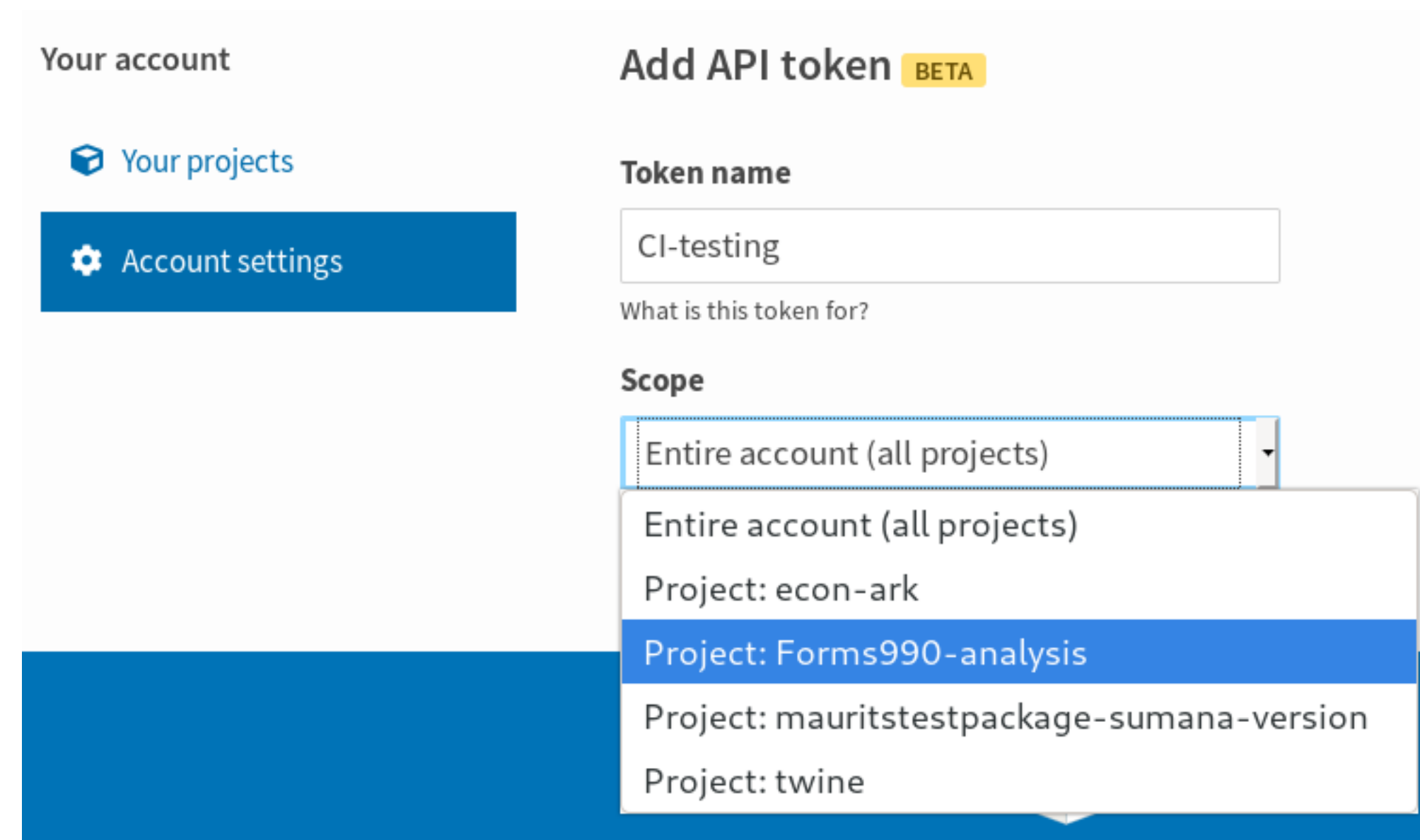
Release: 1.0.0

9 commits later: 1.0.1.dev9+g59ff2f1

- Even accounts for installs from dirty clones



## ► PyPI finally supports API tokens!



The screenshot shows the PyPI account settings interface. On the left, under 'Your account', there are links for 'Your projects' and 'Account settings'. The main content area is titled 'Add API token' with a 'BETA' badge. It contains a 'Token name' input field with the value 'CI-testing', a label 'What is this token for?', and a 'Scope' dropdown menu. The dropdown menu is open, showing a list of options: 'Entire account (all projects)', 'Project: econ-ark', 'Project: Forms990-analysis' (which is highlighted), 'Project: mauritstestpackage-sumana-version', and 'Project: twine'.

<http://pyfound.blogspot.com/2019/07/pypi-now-supports-uploading-via-api.html>



- First: Don't vendor or have builtin dependencies

*How To Make Package Managers Cry: <https://www.youtube.com/watch?v=NSemlYagjIU>*



- First: Don't vendor or have builtin dependencies
- Second: Seriously...don't...

How To Make Package Managers Cry: <https://www.youtube.com/watch?v=NSemlYagjIU>



- First: Don't vendor or have builtin dependencies
- Second: Seriously...don't...
- Third: Make it easy to unvendor them

*How To Make Package Managers Cry: <https://www.youtube.com/watch?v=NSemlYagjIU>*



- First: Don't vendor or have builtin dependencies
- Second: Seriously...don't...
- Third: Make it easy to unvendor them
- Finally: At least make sure they're contained
  - Statically link
  - Ensure symbols are hidden
  - Don't put vendored files in standard locations (e.g. shared libraries in \$PREFIX/lib/)

*How To Make Package Managers Cry: <https://www.youtube.com/watch?v=NSemlYagjIU>*



## ➤ Create Python wheels from conda packages

```
$ conda press --subdir osx-64 --skip-python --fatten iminuit=1.3.7=py37h86efe34_0  
created fat wheel: iminuit-1.3.7-0_py37h86efe34-cp37-cp37m-macosx_10_9_x86_64.whl
```

## ➤ Why?

- Centralises the building of packages
- Easier to pull in dependencies
- Use newer ABIs than manylinux<n>

<https://pydata.org/nyc2019/schedule/presentation/41/conda-press-or-reinventing-the-wheel/>





Questions?