

easybuild



State of the union

Kenneth Hoste (HPC-UGent)

kenneth.hoste@ugent.be

5th EasyBuild User Meeting

Barcelona - Jan 29th 2020



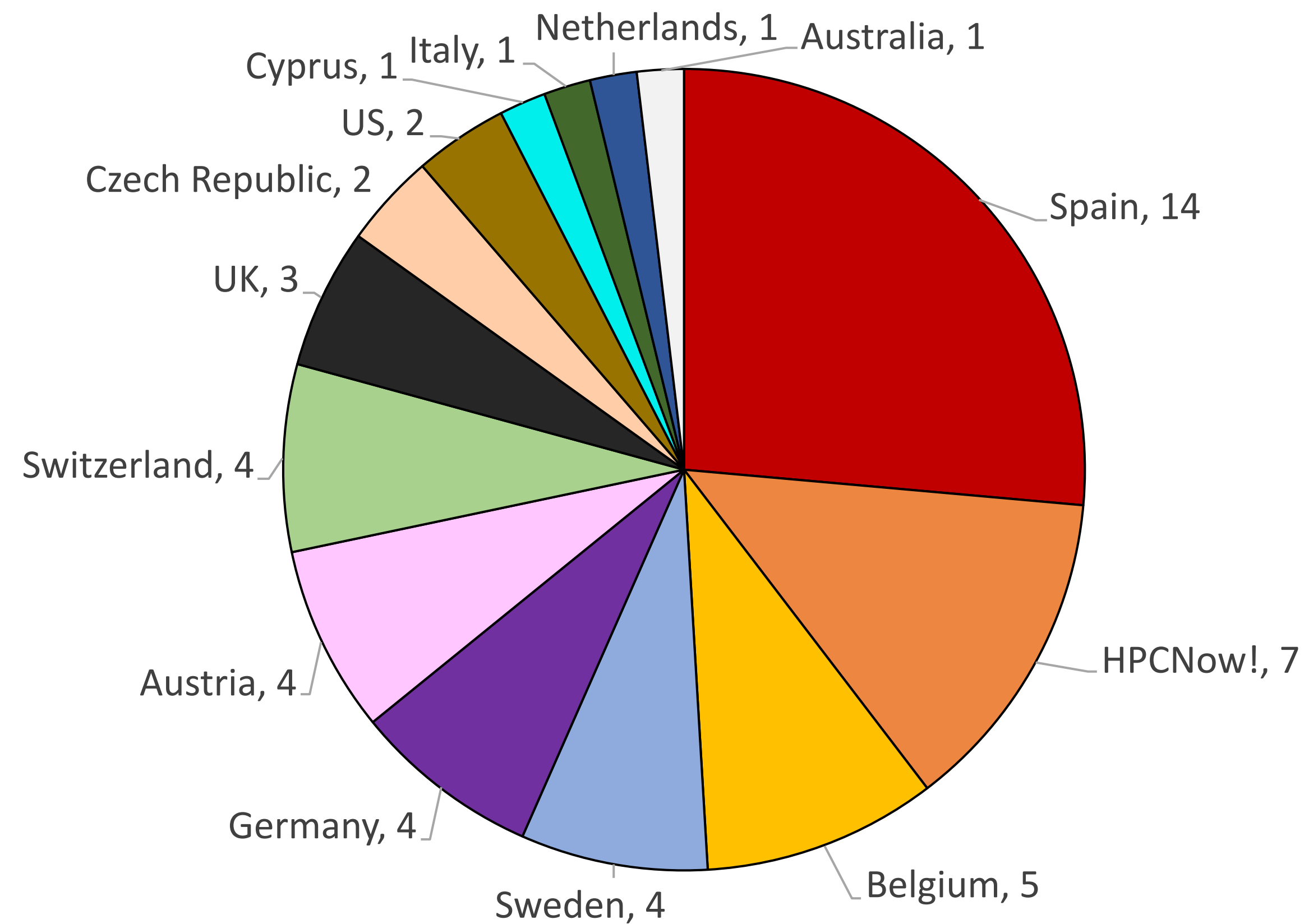
<https://github.com/easybuilders/easybuild/wiki/5th-EasyBuild-User-Meeting>

5th EasyBuild User Meeting - Welcome!

see also <https://tiny.cc/eum20>

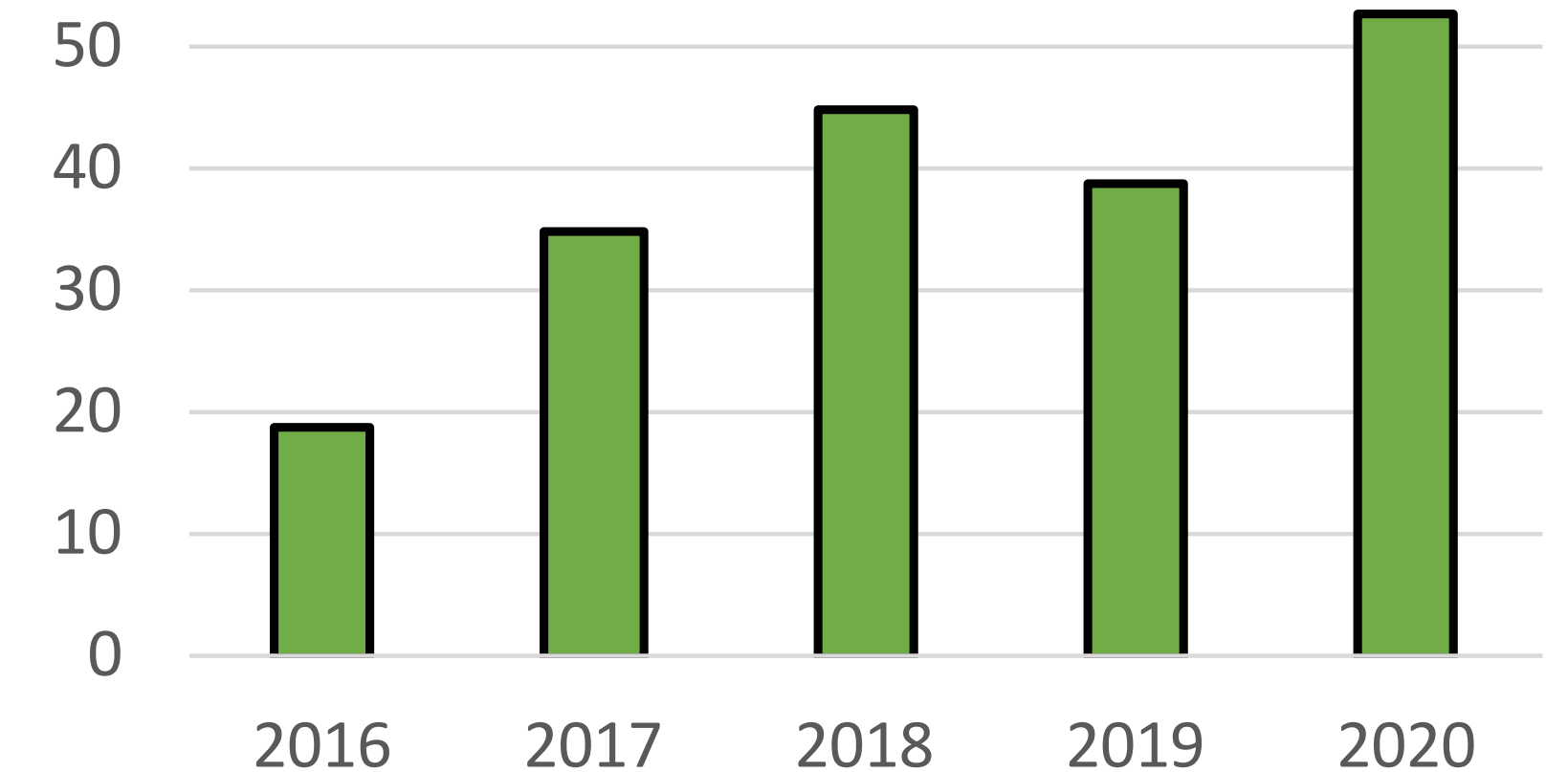


5th EasyBuild User Meeting: attendance



EUM attendees by country (in 2020)

13 countries, 3 continents



Attendance is growing over the years:

- 2016: 19 (HPC-UGent, Belgium)
- 2017: 35 (JSC, Germany)
- 2018: 45 (SurfSARA, Netherlands)
- 2019: 39 (CECI, Belgium)
- **2020: 53 (HPCNow!, Spain)**

5th EasyBuild User Meeting - agenda

Wed Jan 29th 2020

see also <https://tiny.cc/eum20>



State of the Union
Kenneth Hoste



**Providing a Unified User Environment for
Canada's National Computing Centers**
Maxime Boissonneault (remote)



John Dey

FRED HUTCH™



**Building an EasyBuild
Container Library
in Sylabs Cloud**
Shahzeb Siddiqui



Massimiliano Culpo

5th EasyBuild User Meeting - agenda

Wed Jan 29th 2020

see also <https://tiny.cc/eum20>



[20:00] informal group dinner @ La Esquinica (tapas)
each at own cost

If you want to join, register & pay ASAP!

5th EasyBuild User Meeting - agenda

Thu Jan 30th 2020

see also <https://tiny.cc/eum20>

buildtest

Shahzeb Siddiqui

ReFrame

Vasileios Karakasis

openstack®

Ümit Seren Vienna
BioCenter



Contributing back to



Samuel Moors

Lmod



Robert McLay (remote)

TACC

site presentations



Oriol Mula



Luca Marsella

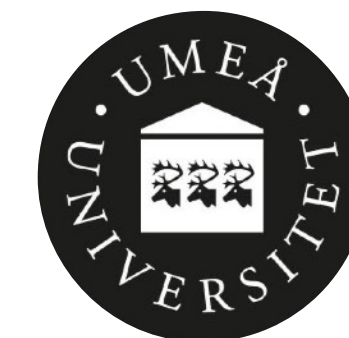
BEAR

BIRMINGHAM ENVIRONMENT
FOR ACADEMIC RESEARCH

Simon Branford



*Caspar
Van Leeuwen*



Åke Sandgren

5th EasyBuild User Meeting - agenda

Thu Jan 30th 2020

see also <https://tiny.cc/eum20>



[20:30] group dinner @ Can Cortada
sponsored by HPC-UGent

5th EasyBuild User Meeting - agenda

Fri Jan 31st 2020

see also <https://tiny.cc/eum20>



Visit to Barcelona Supercomputing Center

bus leaves at:

- ***7.30am (hotel Ibis Meridiana)***
- ***7.45am (Parc Tecnològic)***

**Sarus: Highly Scalable Docker Containers
for HPC Systems**
Luca Marsella

site presentations



Jin Zhang

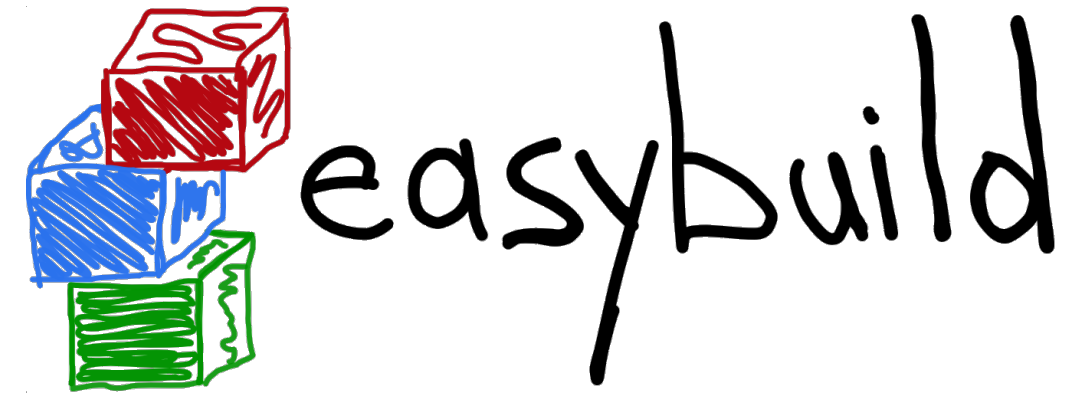


Alexander Grund

**10 things you didn't know yet about
Kenneth Hoste**



Live streaming & recording of all EUM20 talks



- Live streaming of all talks via the **EasyBuild YouTube channel**
<https://www.youtube.com/c/easybuilders>
- Talk recordings should be available there shortly after each talk
- Talks from all 4 previous EasyBuild User Meetings are also listed there, together with other talks on EasyBuild and related tools (see playlists)
- Join #eum channel in EasyBuild Slack for Q&A!
- **Many thanks to Alan O'Cais (JSC) for the help with this endeavour!**



whoami

kenneth.hoste@ugent.be
@boegel (*GitHub, IRC, Slack*)
@kehoste (*Twitter*)

- Masters & PhD in Computer Science from Ghent University
- PhD topic: machine learning applied to software performance, compilers, ...
- joined HPC-UGent team in October 2010
- main tasks: user support & training, *software installations*
- slowly also became  *easybuild* **lead developer (& release manager)**
- likes family, beer, loud music, FOSS, helping people, dad jokes, stickers, ...
- doesn't like CMake, SCons, Bazel, TensorFlow, OpenFOAM, ...

HPC-UGent

<https://www.ugent.be/hpc/en>




- part of central IT department of Ghent University (Belgium)
- centralised scientific computing services, training & support
- for researchers of UGent, industry & knowledge institutes
- 6 Tier-2 clusters (> 14k cores in total), ~2 PB shared storage
- 8+1 team members, ~3.5k user accounts
- member of Flemish Supercomputer Centre (VSC)

<https://www.vscentrum.be>

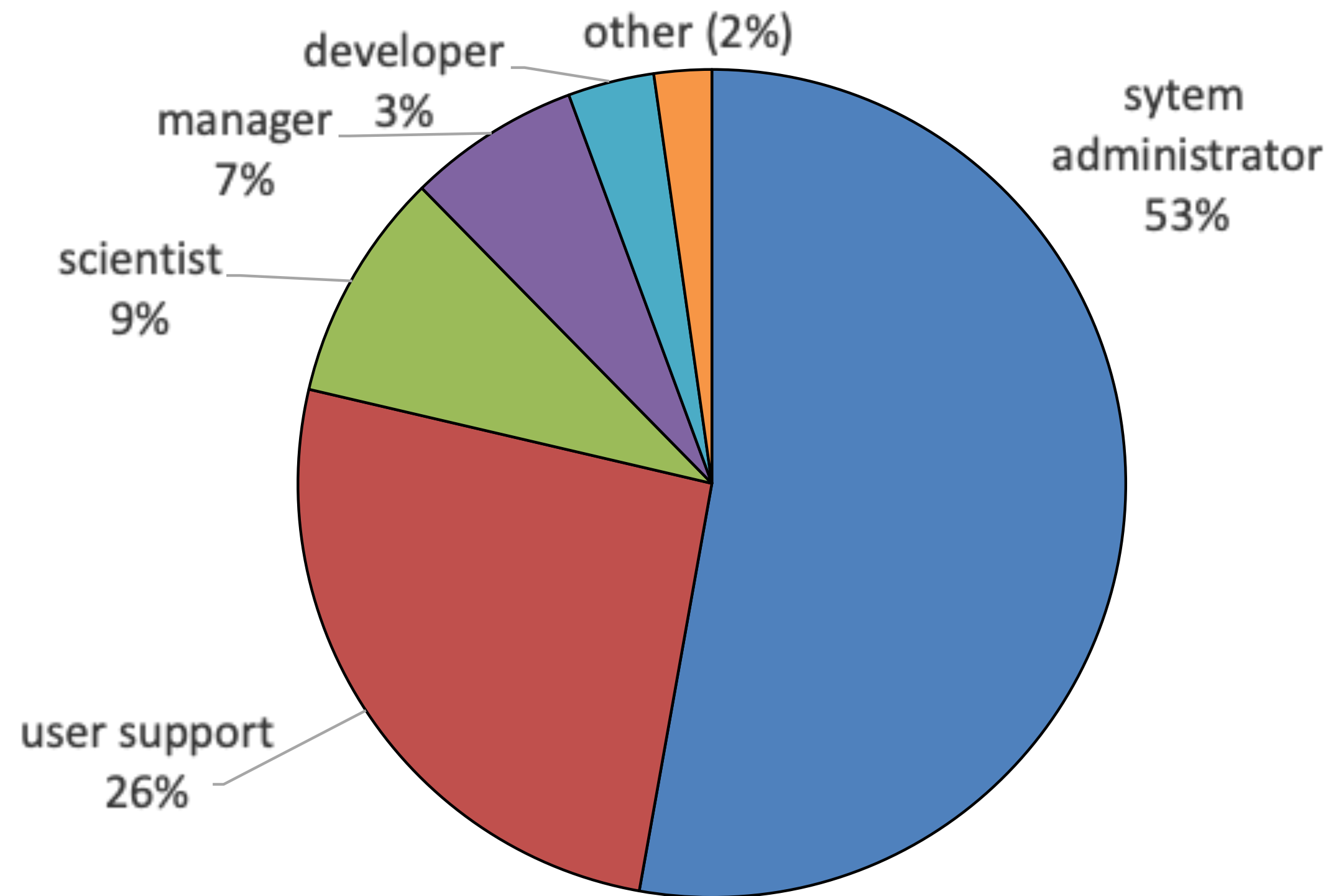




easybuild User Survey

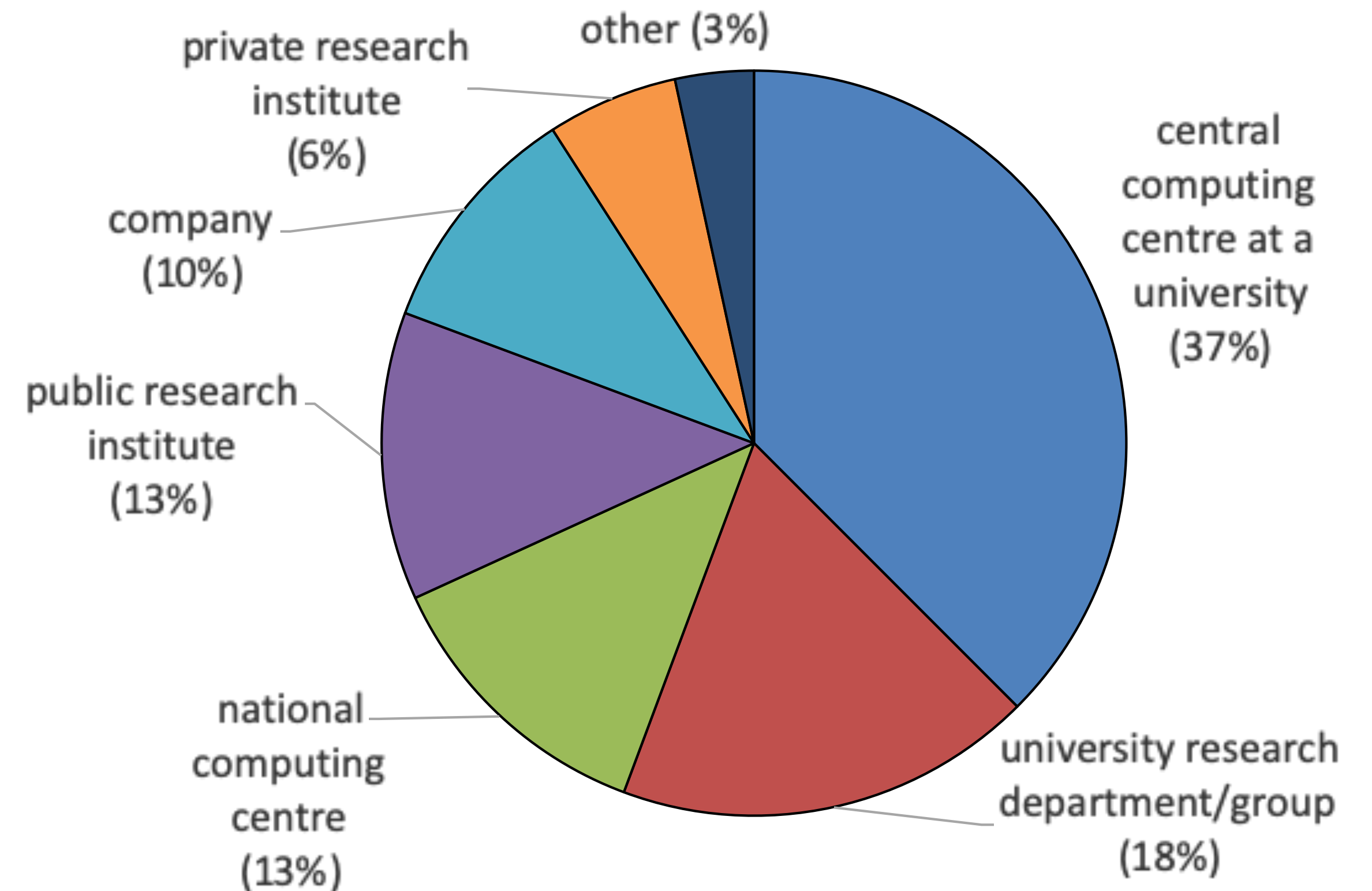
- anonymous survey via  **SurveyMonkey**[®]
- main objective is to get feedback from / insight into EasyBuild community
- 39 questions, covers wide variety of aspects
- 3rd year in a row (also for EUM'18 & EUM'19)
- 91 people participated (77 in 2018, 93 in 2019)
- assumed to give relevant view on community
(but results are interpreted with a grain of salt)
- results are discussed in this presentation (out of order)

What is your primary profile?



Clear bias towards sysadmins & user support.

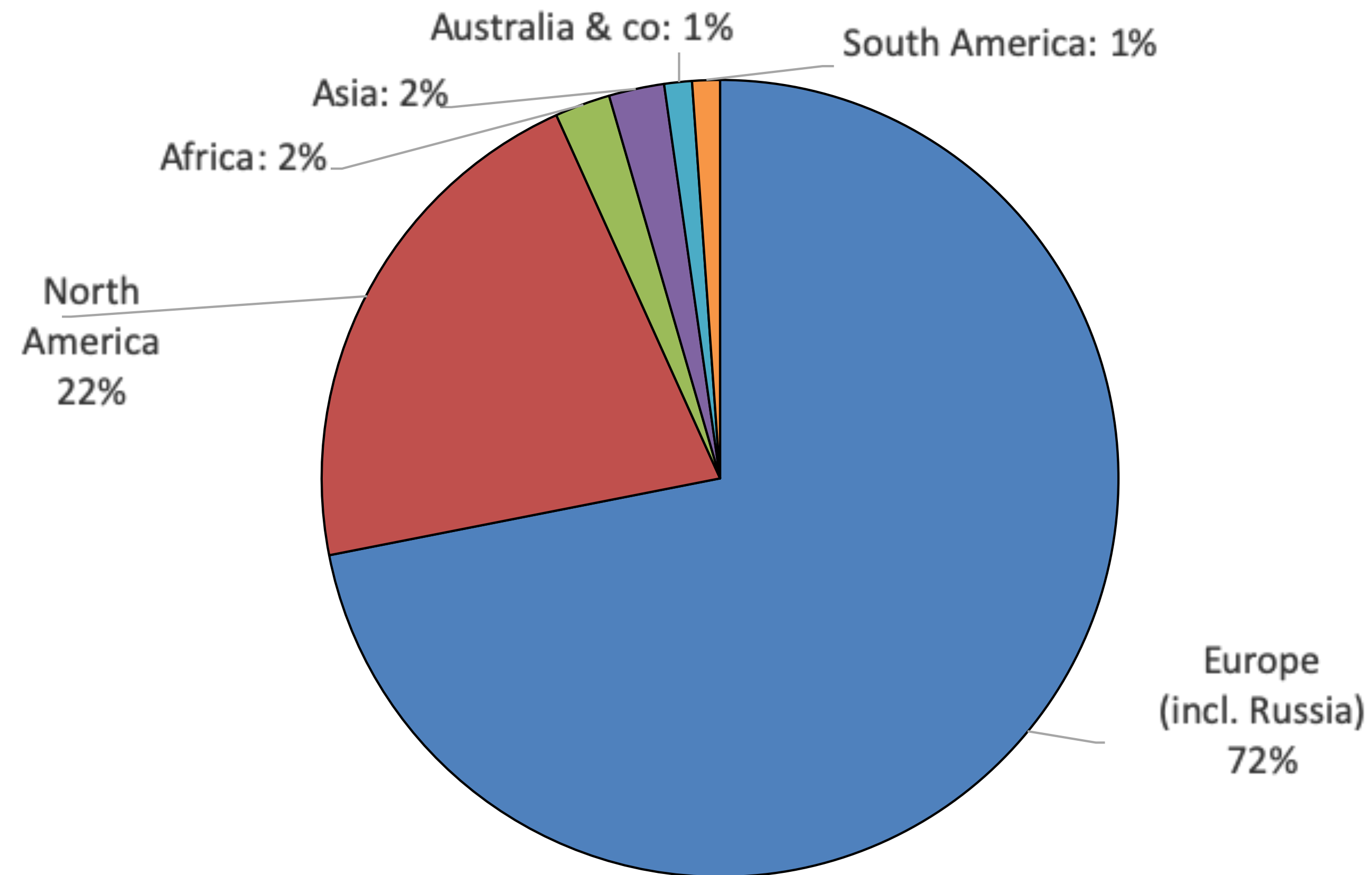
What type of organisation do you work for?



(very similar results compared to previous 2 surveys)

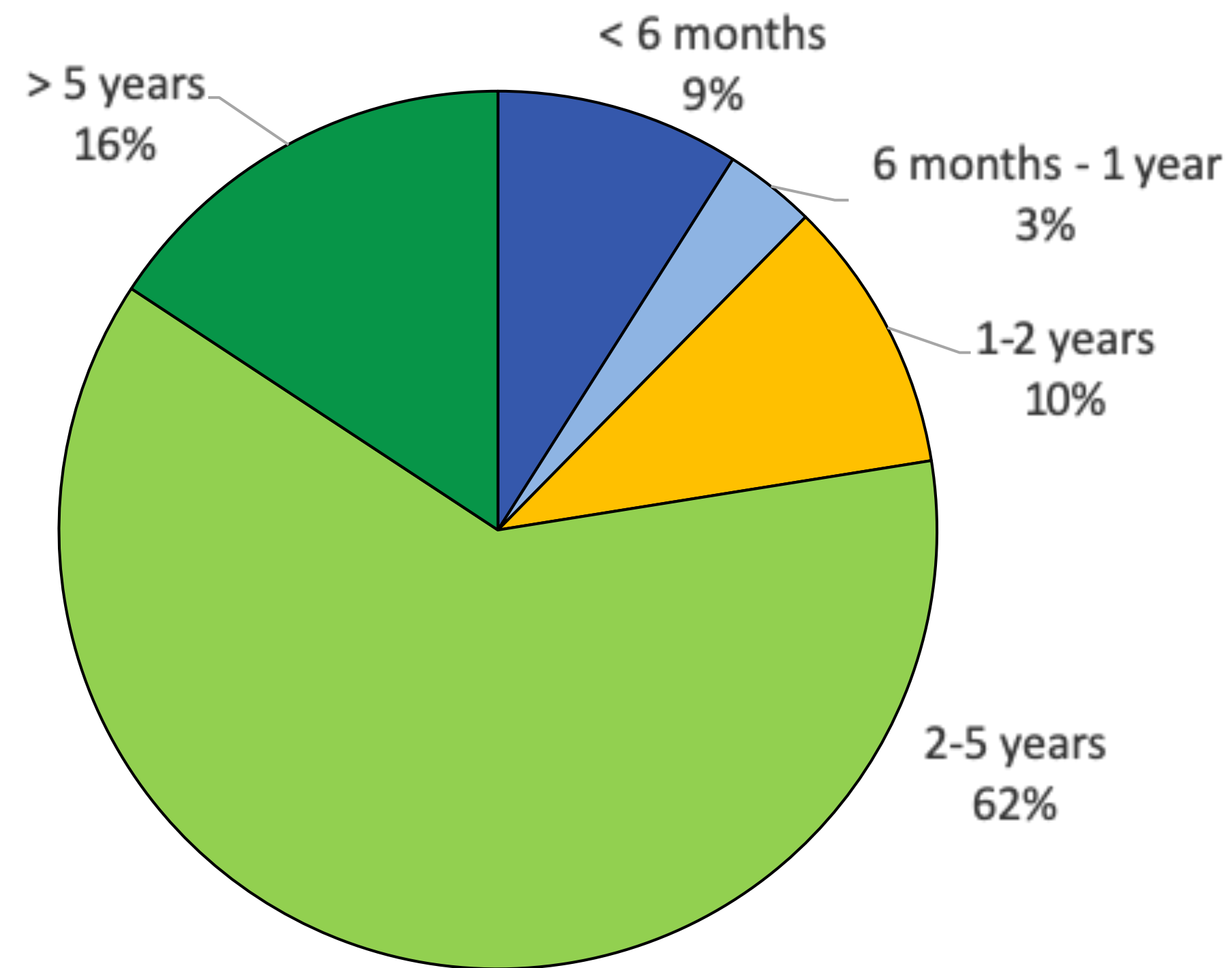
3rd easybuild User Survey

In which part of the world are you located?



(very similar results to previous survey)

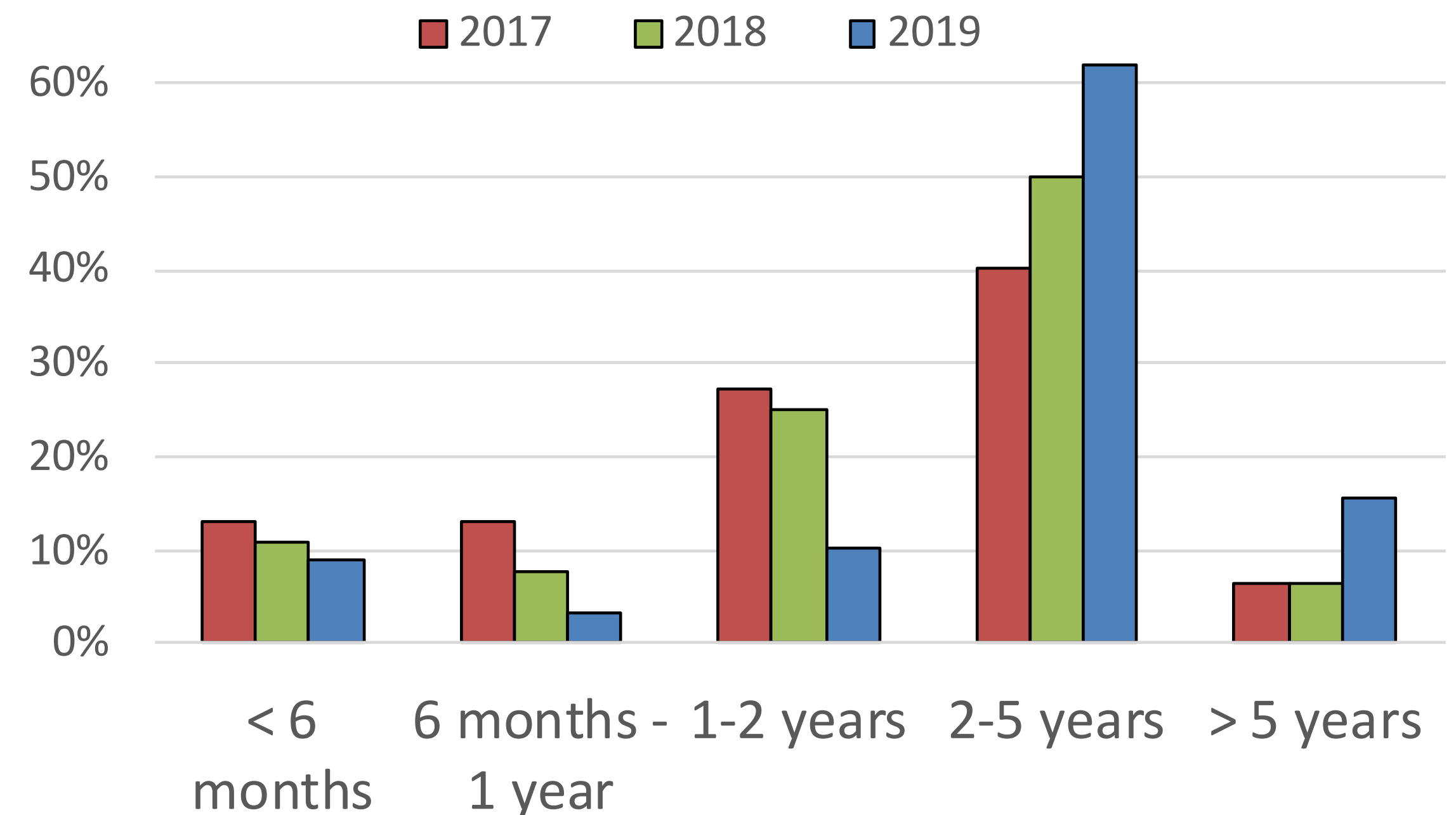
How long have you been using EasyBuild?



results of this survey (2019)

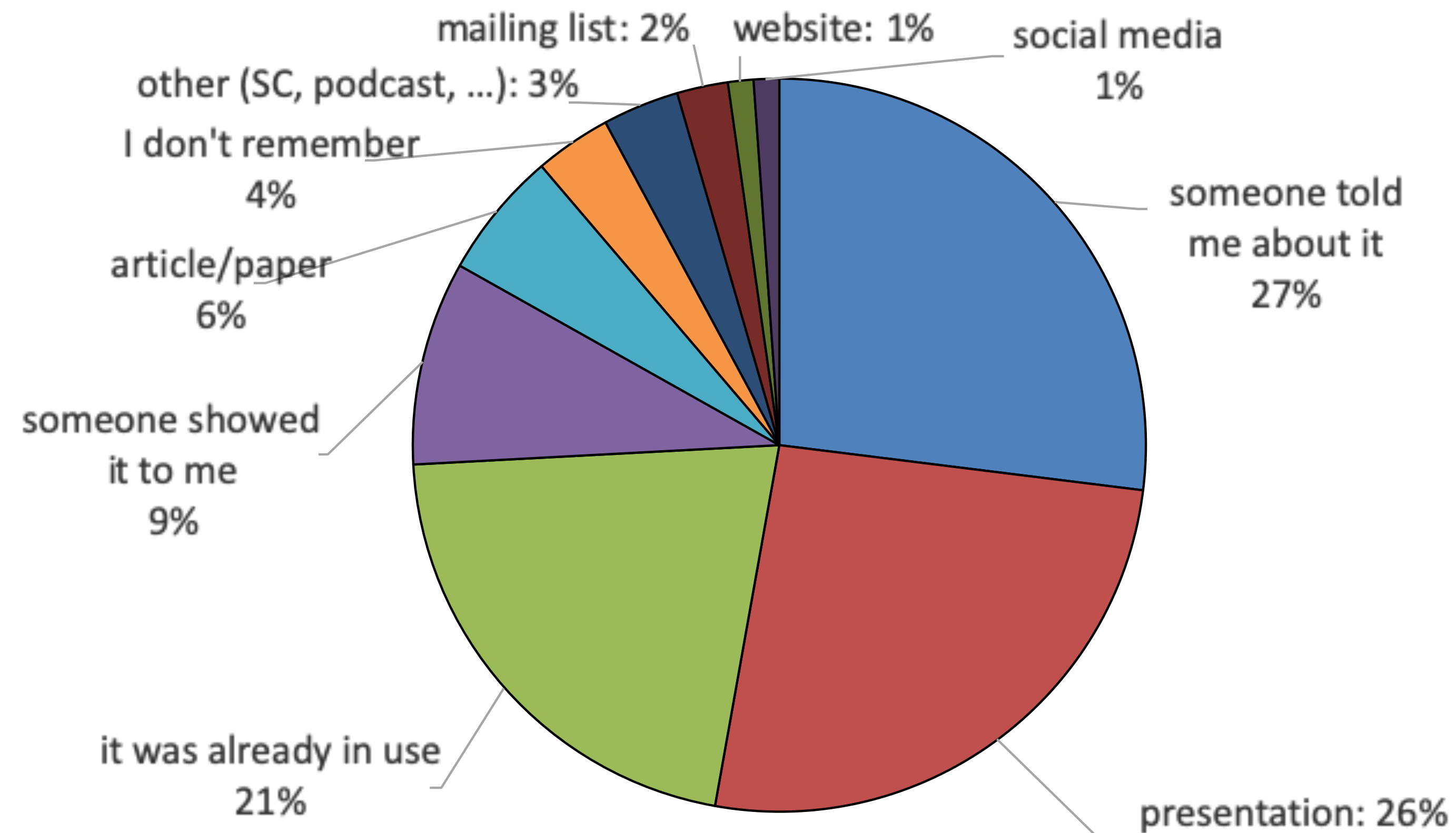
Evolution over the years:

- shift to long-time users
- new users keep coming in
- sites that start using EasyBuild stick with it?



3rd easybuild User Survey

How did you first learn about EasyBuild?

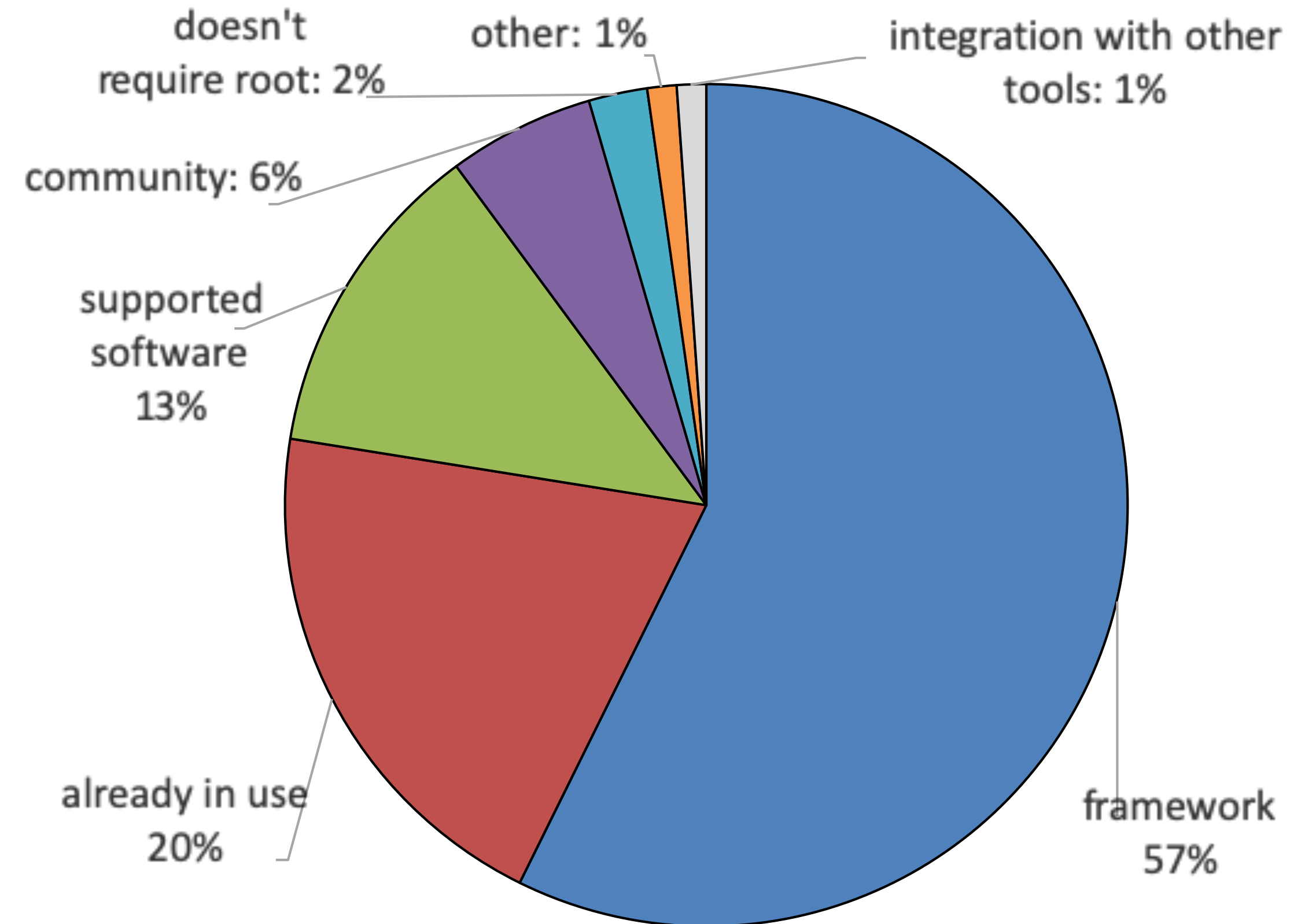


Majority is still by word-of-mouth,

"it was already in use" keeps increasing (4% ➡ 19% ➡ 21%)

3rd easybuild User Survey

What is the main aspect of EasyBuild that convinced you to start using it?



Core functionality & supported software remain most convincing.

Highlights of 2019



- **EasyBuild 4.0 release** \o/
- (continued) growth of EasyBuild community
- (continued) growth in contributions (and managing to keep up with them)
- 12% increase in # supported software packages (despite archiving old stuff)
- better support for Python packages (PythonBundle, multi_deps, pip check, etc.)
- (also) using GitHub Actions for CI tests => test results are available sooner
- improved GitHub integration
- improved support for building (Singularity) container images

Important changes in easybuild v4.0

- compatible with both Python 2 (2.6, 2.7) & **Python 3** (3.5+)
- no more required Python packages (Python standard library is sufficient)
 - vsc-base & setuptools are no longer a dependency
- deprecated dummy toolchain, replaced by `system` toolchain
- detection of unknown easyconfig parameters being used
 - + recommended naming scheme for local variables in easyconfig files

More information:

<https://easybuild.readthedocs.io/en/latest/EasyBuild4-overview-of-changes.html>

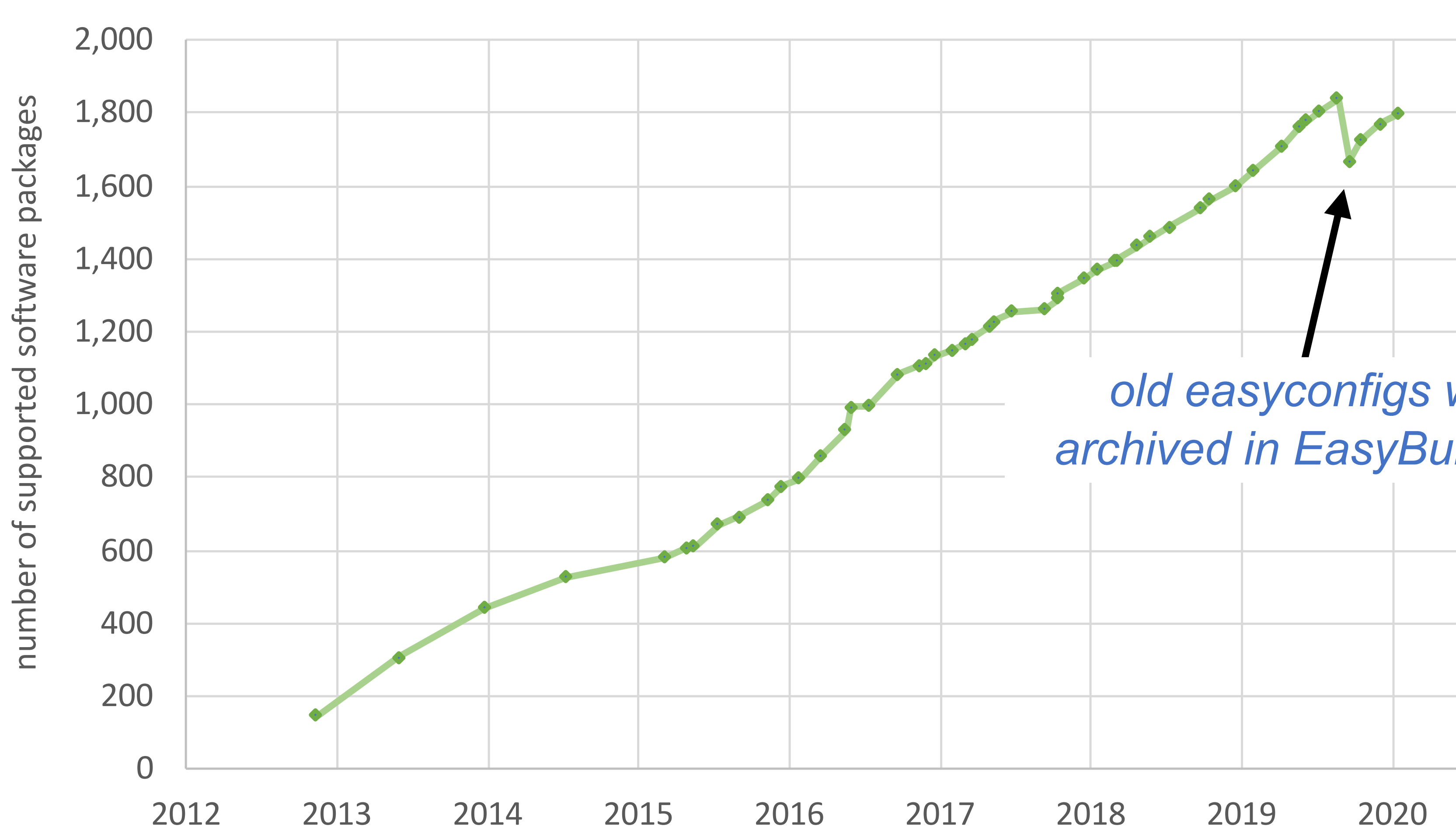
The road to Python 3



- effort to port EasyBuild to Python 3 started in Dec'18
- step 0 was to ingest code from vsc-base & drop dependency (Dec'18)
- first goal: make all framework tests pass in Python 3.6 (done by end of Feb'19)
- also test with Python 3.5 and 3.7 (done by mid March'19)
- testing of actual software installations + various bug fixes (March-Sept)
- **EasyBuild v4.0 release on Sept 20th 2019**
- a couple of (small) bug fixes in following weeks & months
- since EasyBuild v4.1.1: no known problems when running on top of Python 3

Supported software

Evolution of number of supported software packages (without extensions)

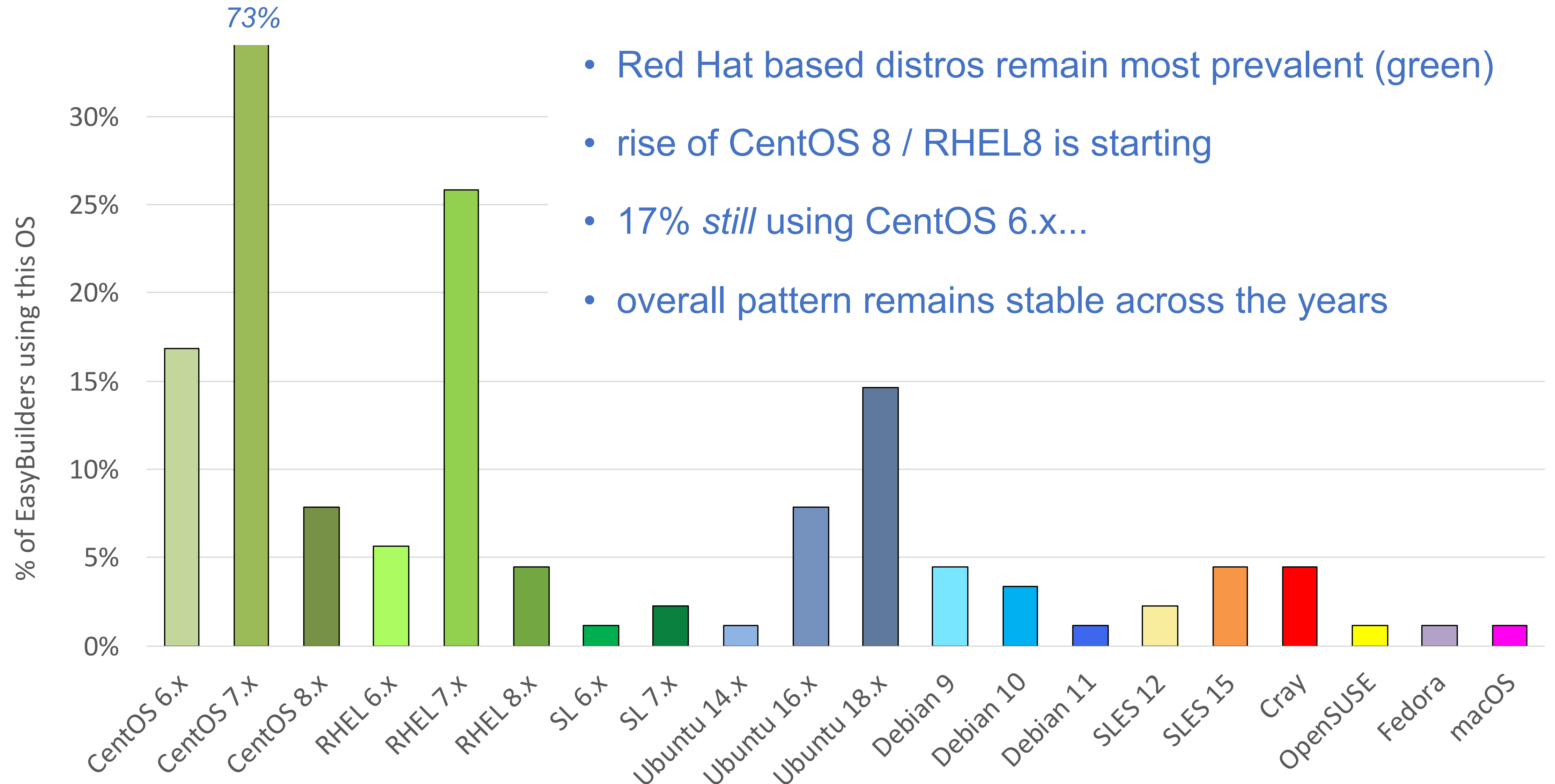


- 1/3rd of supported software is bioinformatics
- 23% is (generic) libraries & tools
- > 2000 supported software in 2020?

3rd easybuild User Survey

On which operating system(s) do you use EasyBuild (most commonly)?

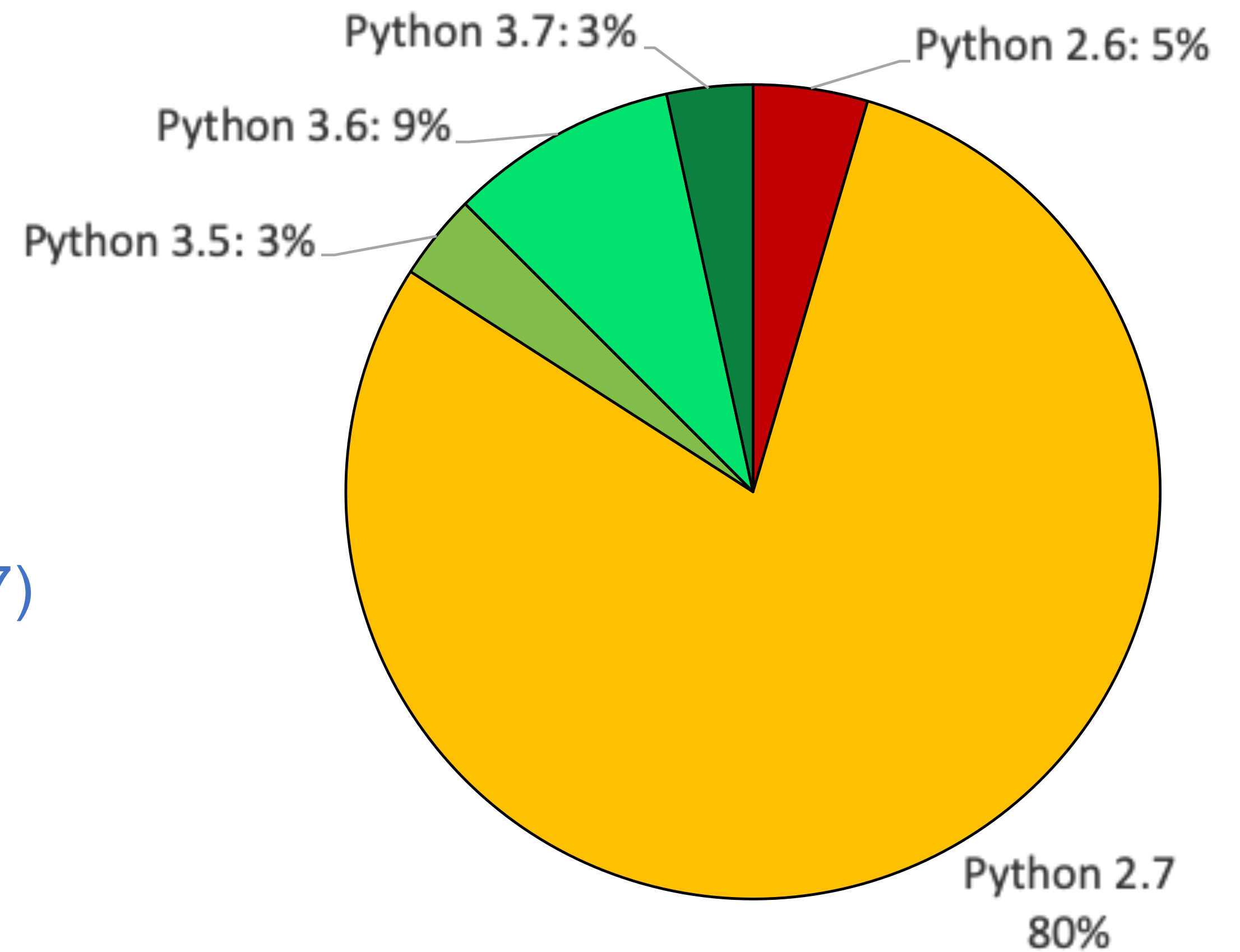
(multiple answers possible)



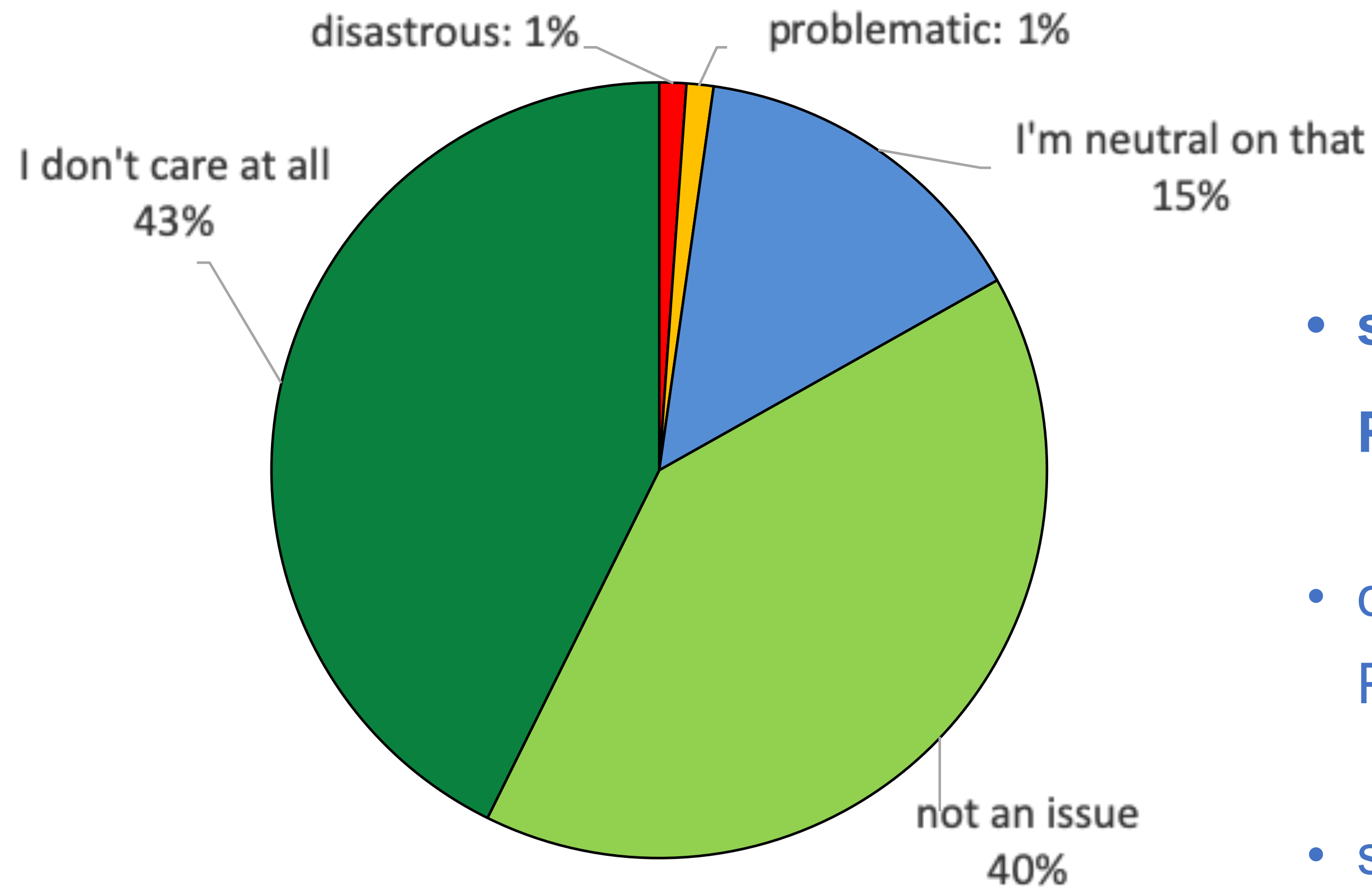
3rd easybuild User Survey

Which Python version do you usually use to run EasyBuild?

- vast majority (still) uses Python 2.7 (84% ➡ 80%)
- small minority still using Python 2.6 (CentOS 6?)
- splintered use across different Python 3 versions
- my predictions for 2020 (next survey):
 - majority still using Python 2 (CentOS 7/RHEL 7)
 - significant increase in use of Python 3
 - Python 3.6 most prevalent Python 3 version

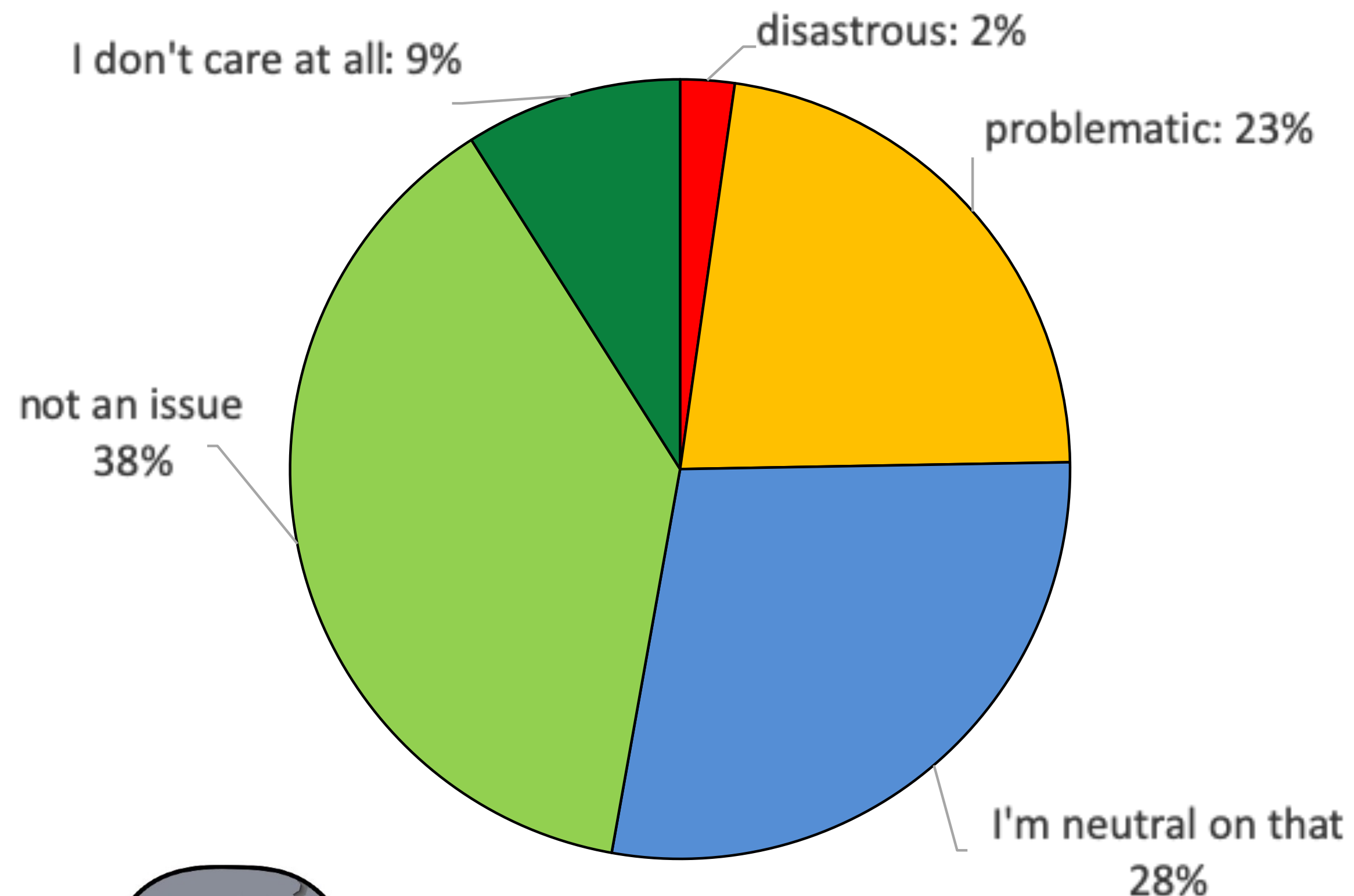


How troublesome would it be for you if EasyBuild becomes incompatible with Python 2.6?



- support for running EasyBuild on top of Python 2.6 was deprecated in EasyBuild 4.0
- declining number of people care about Python 2.6 (13% ➡ 4% ➡ 2%)
- some people stuck on CentOS 6 or RHEL 6 have installed Python 2.7

How troublesome would it be for you if EasyBuild would be only compatible with Python 3?



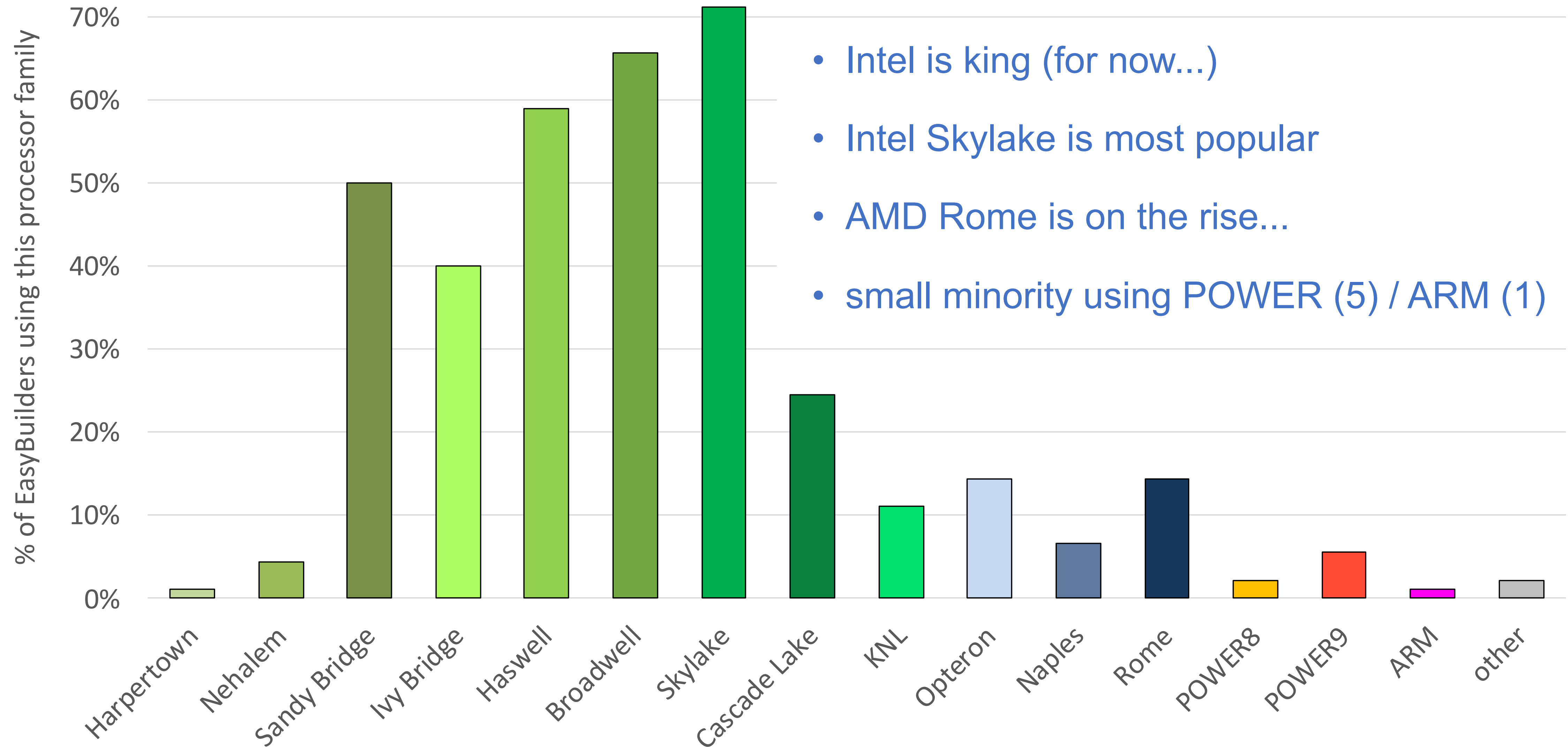
- ~25% is concerned about dropping support for Python 2
- **no plans currently to drop Python 2 support**
- keeping Python 2 supported is not a big issue...
- no significant benefit in only supporting Python 3?
- likely scenario:
 - deprecate Python 2 support in some future EasyBuild version (4.x)
 - stop supporting Python 2 in EasyBuild 5.0 (?)



3rd easybuild User Survey

For which processor families are you installing software using EasyBuild?

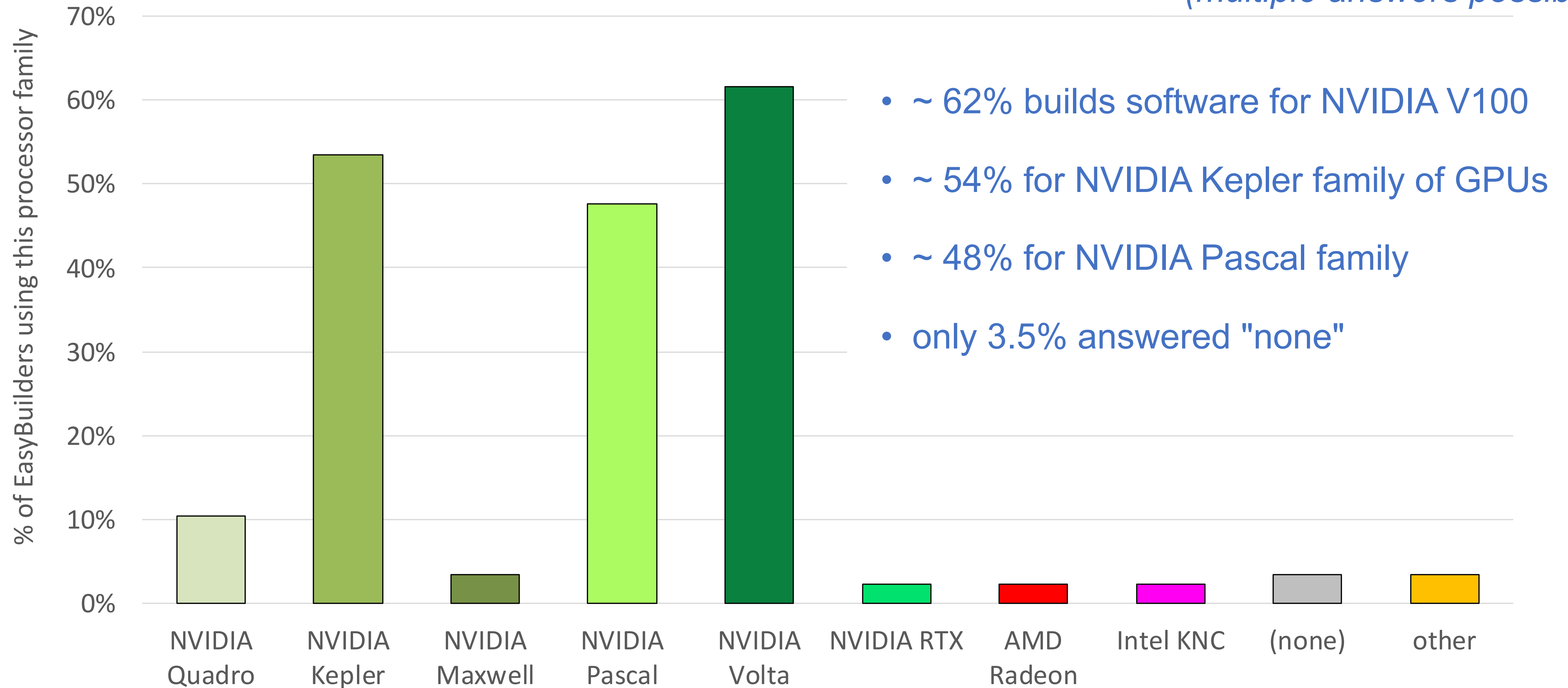
(multiple answers possible)



3rd easybuild User Survey

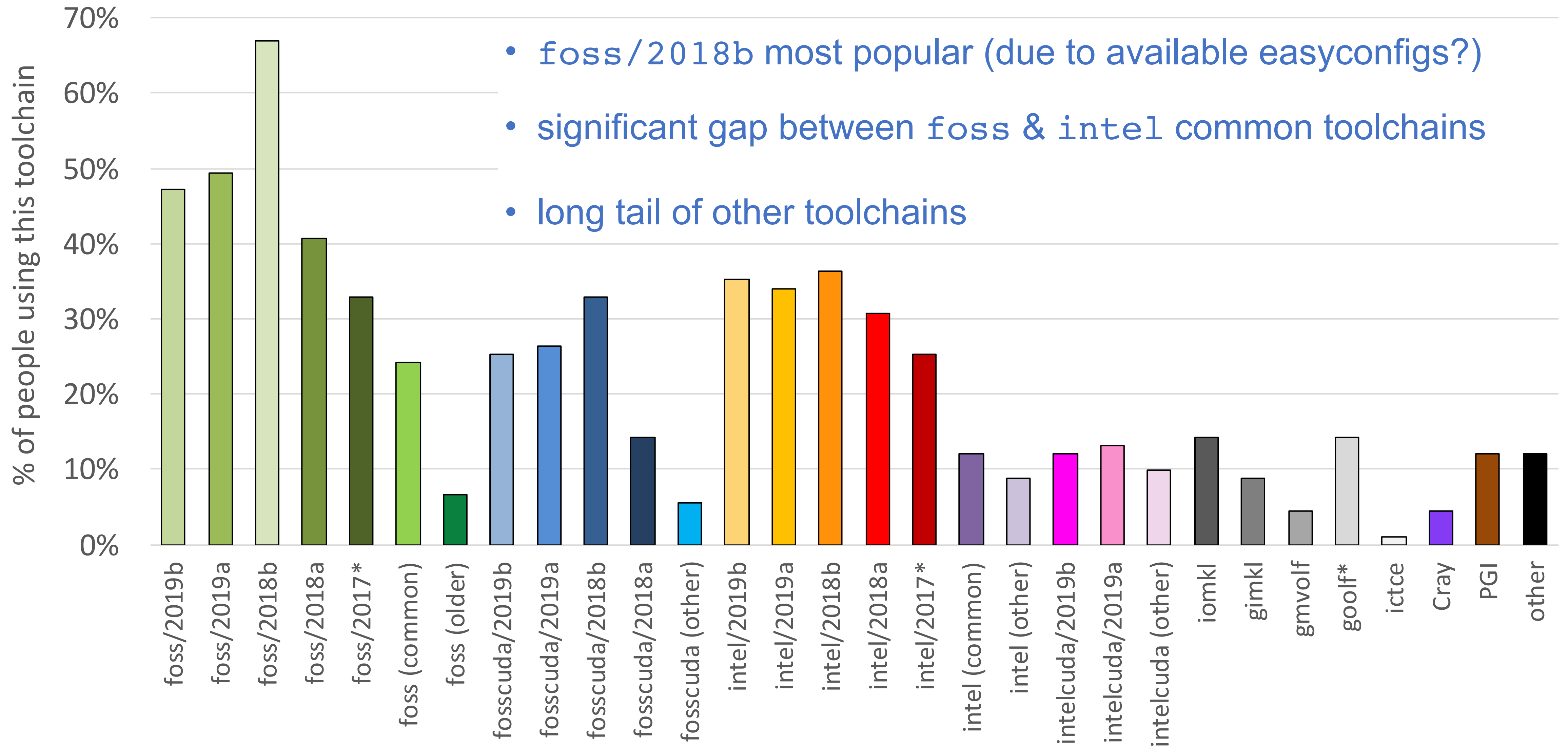
For which accelerators are you installing software using EasyBuild?

(multiple answers possible)

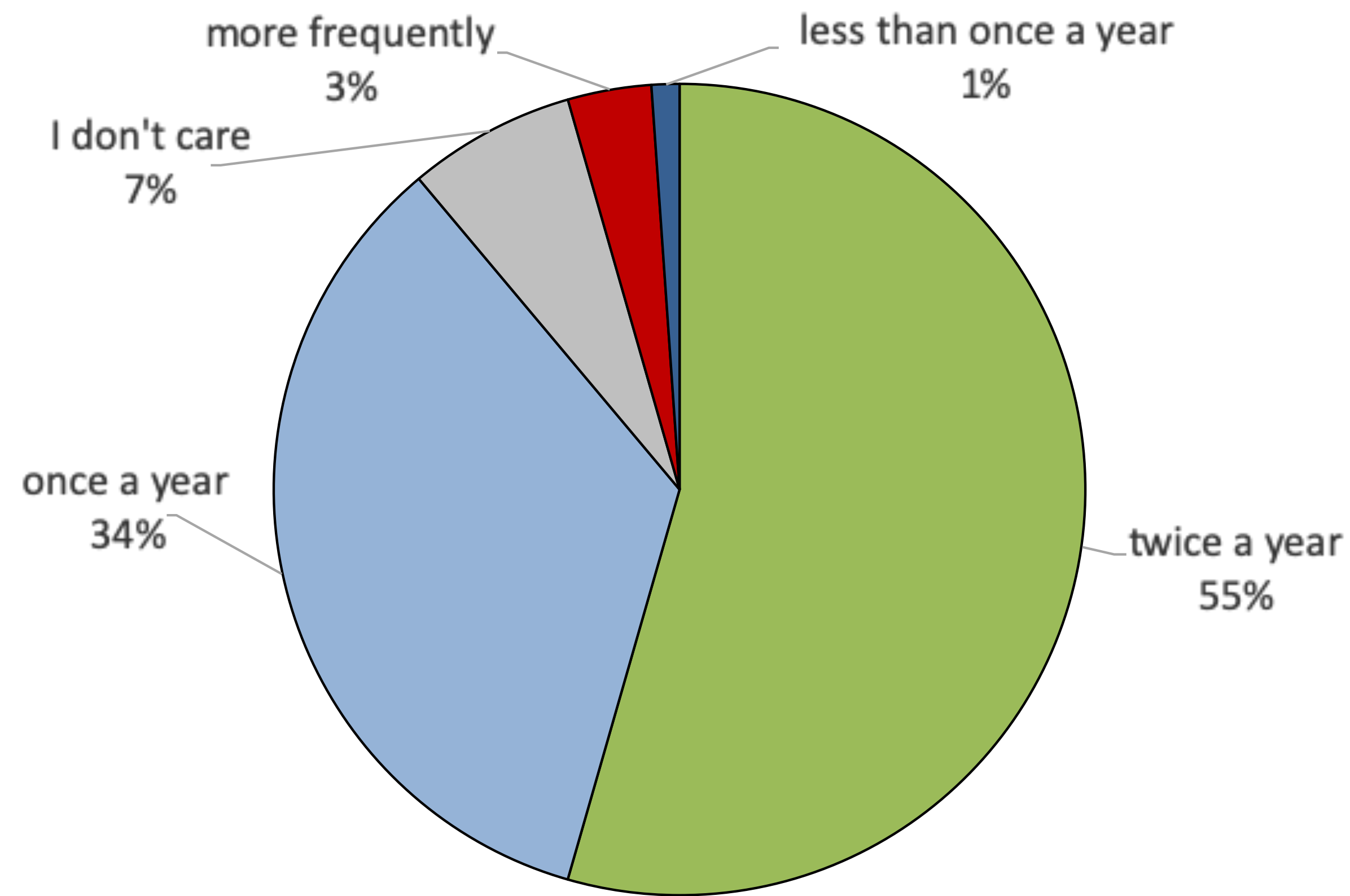


3rd easybuild User Survey

Which (full) toolchain(s) do you use? (check all that apply)



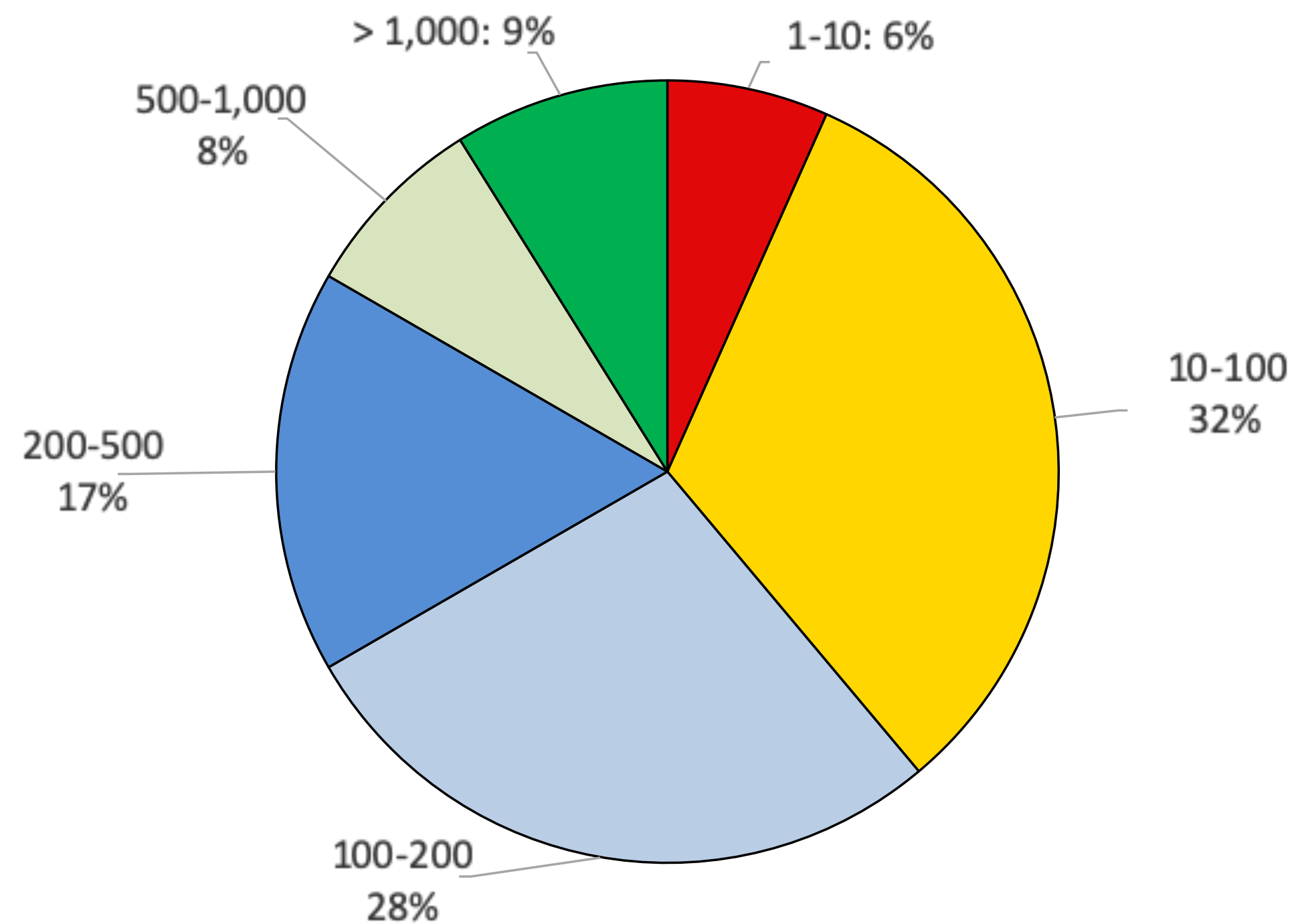
How frequently should the 'foss' and 'intel' common toolchains be updated?



- preference for current update frequency (twice per year) remains stable (57% ➡ 55%)
- slight increase for "once a year" (28% ➡ 34%)
- no changes planned (but willing to reconsider)

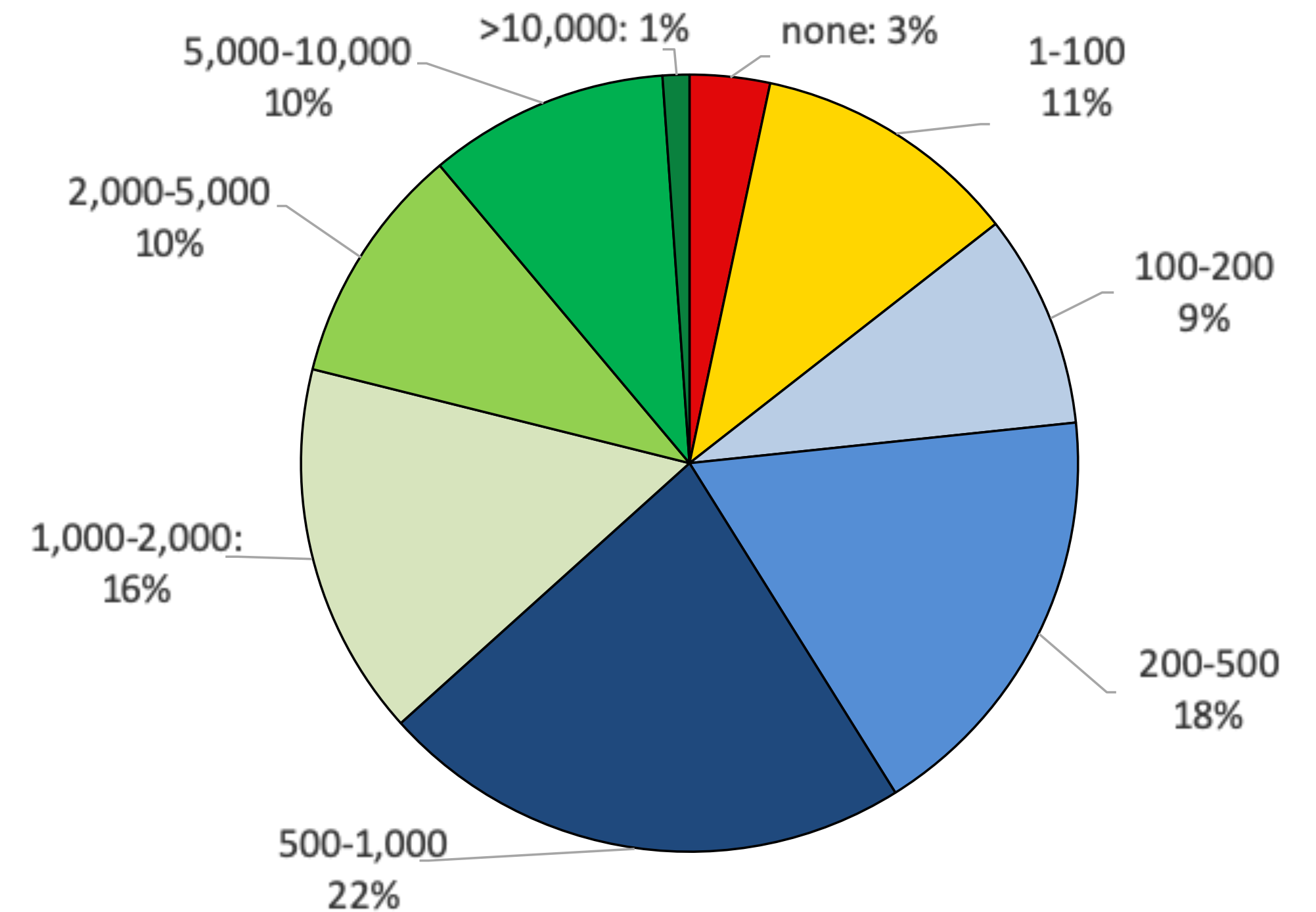
3rd easybuild User Survey

How many software installations did you perform in the last year using EasyBuild?



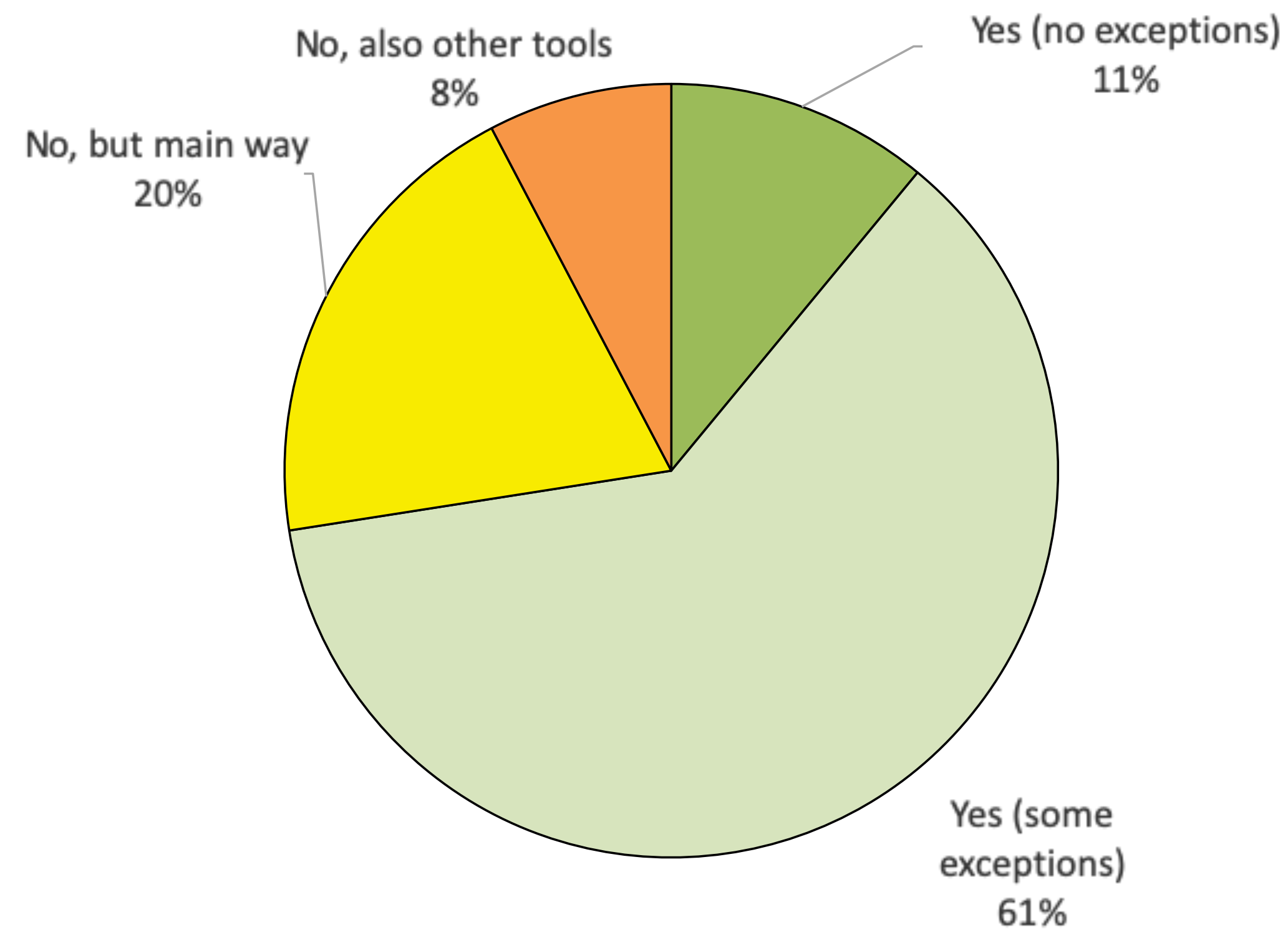
small shift to > 100 installations / year

How many software installations/modules installed with EasyBuild do you currently have in production?



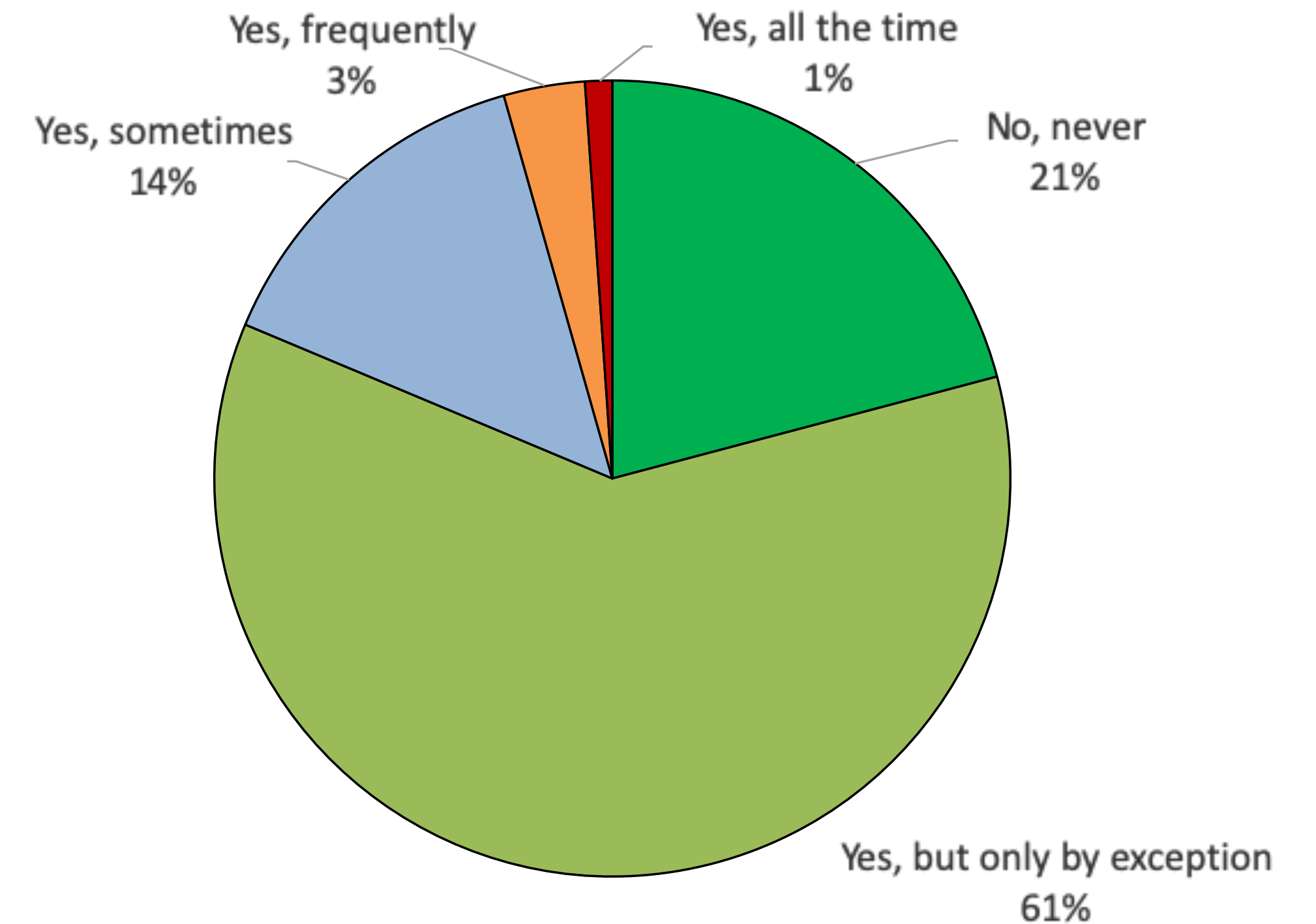
(new question)

Is EasyBuild your only way of installing (scientific) software?



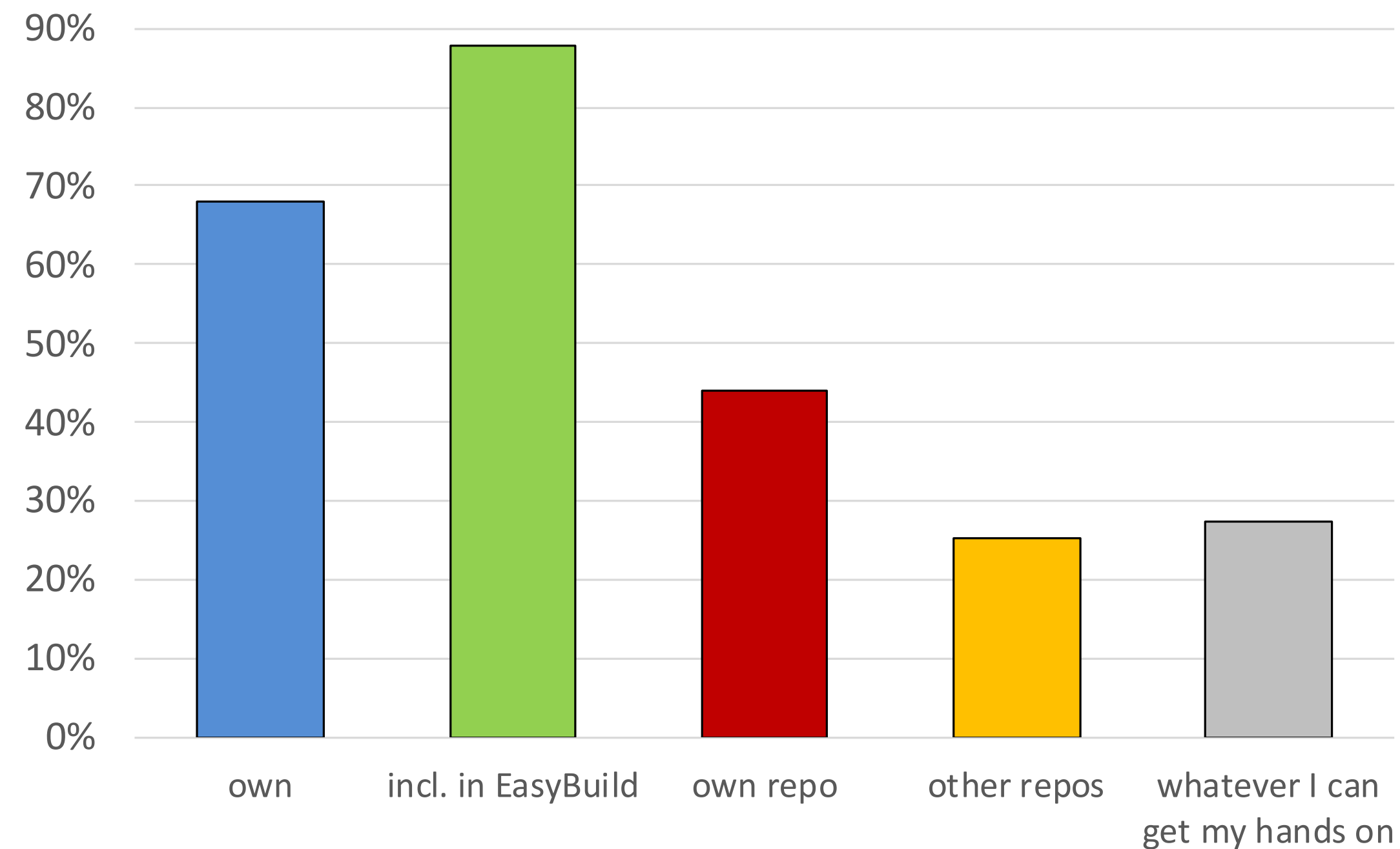
small increase in "yes" (69% ➡ 72%)

Do you still install (scientific) software manually?



small increase in "never" (15% ➡ 21%),
decrease in "sometimes" (23% ➡ 14%)

Which easyconfig files do you use? (check all that apply)



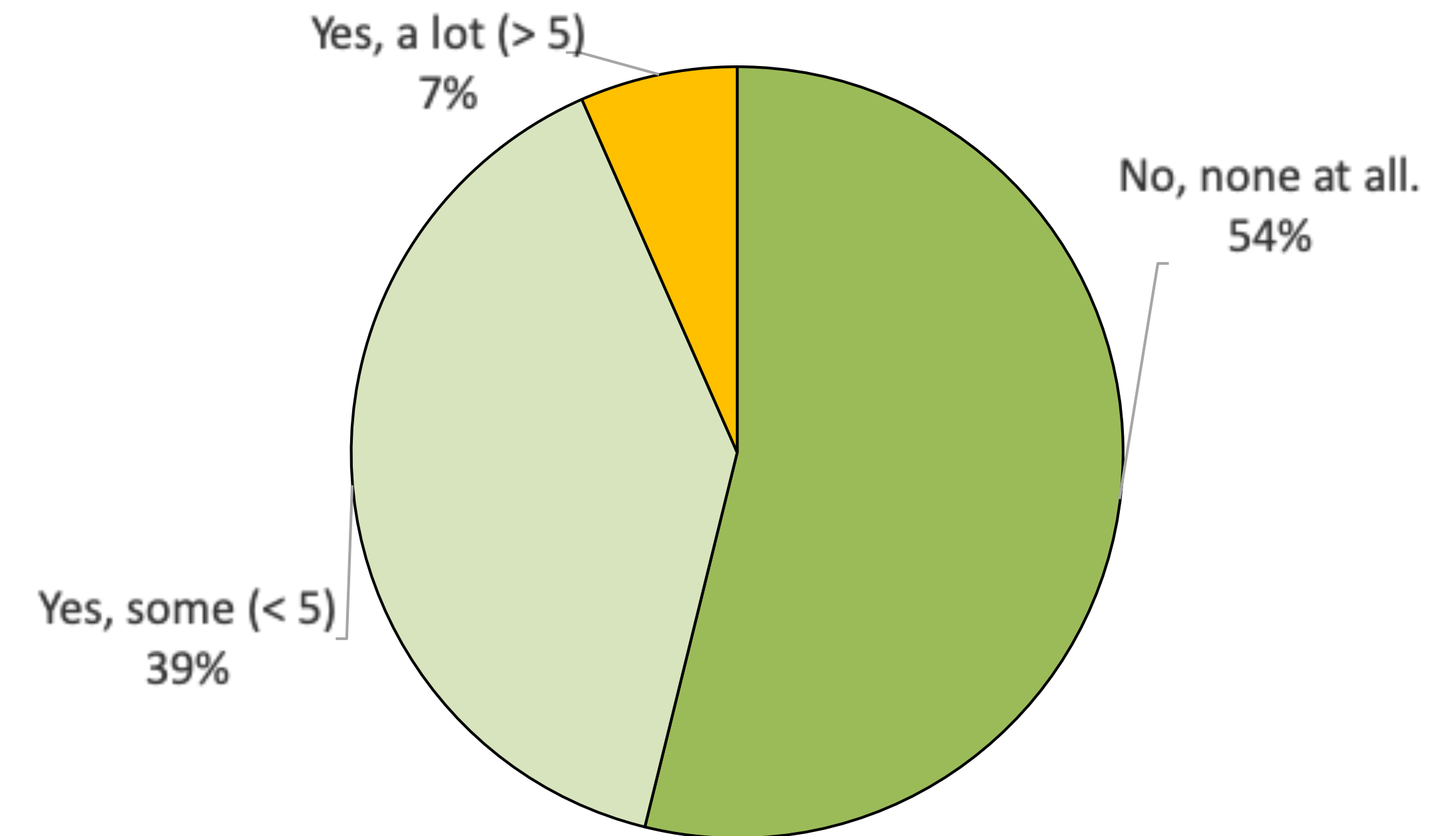
88% uses easyconfigs included with EasyBuild

68% uses own easyconfigs

44% maintains own repo

(very similar to previous survey)

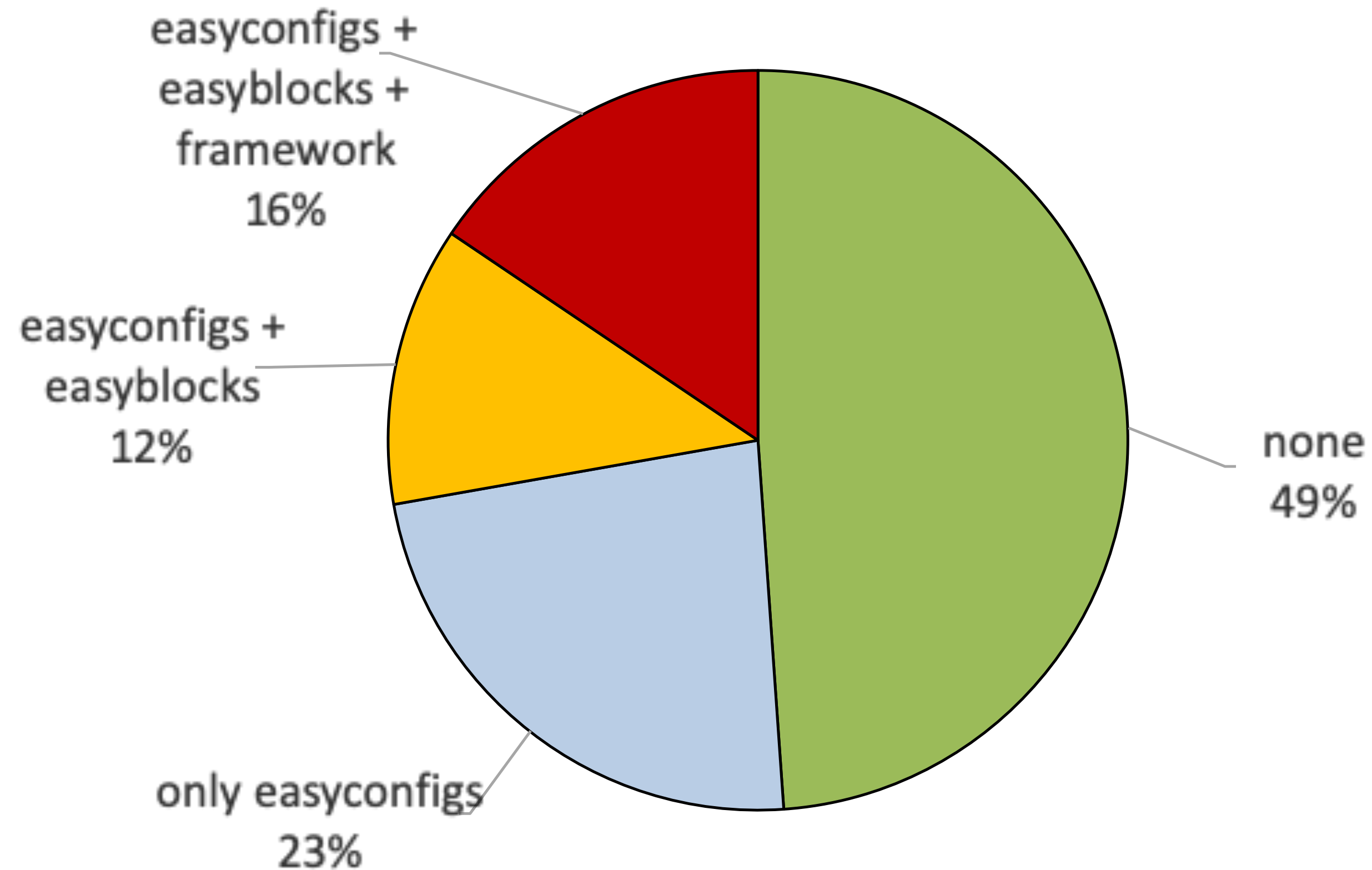
Do you use any custom easyblocks?



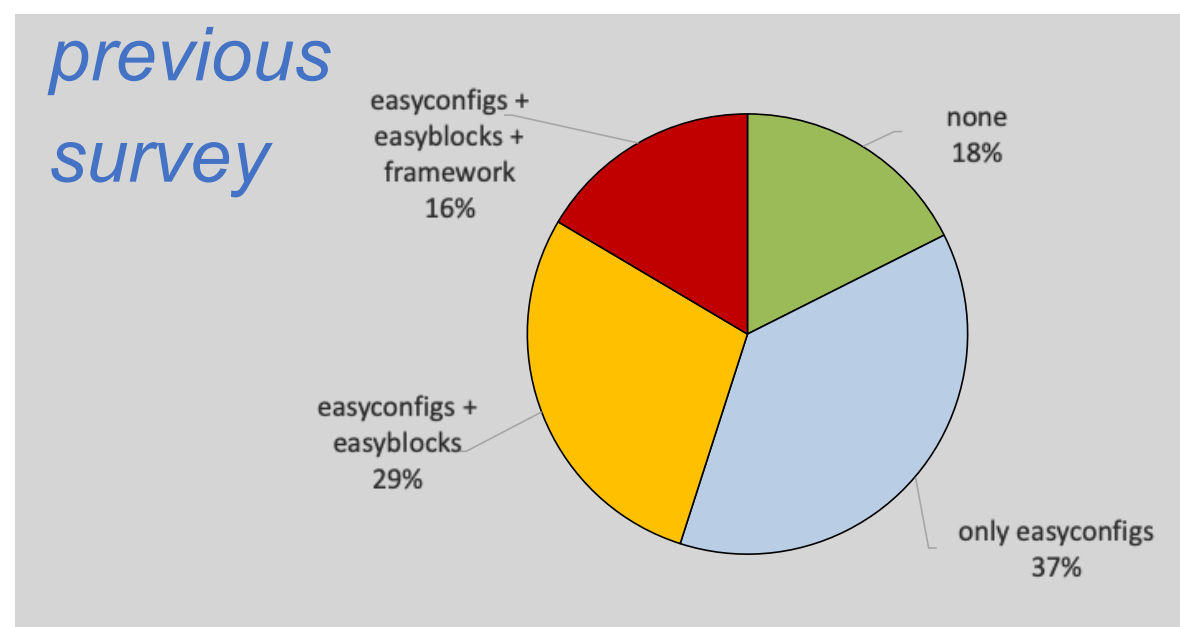
small increase in "None" (49% ➡ 54%)

3rd easybuild User Survey

Have you made any site-specific customisations to EasyBuild?

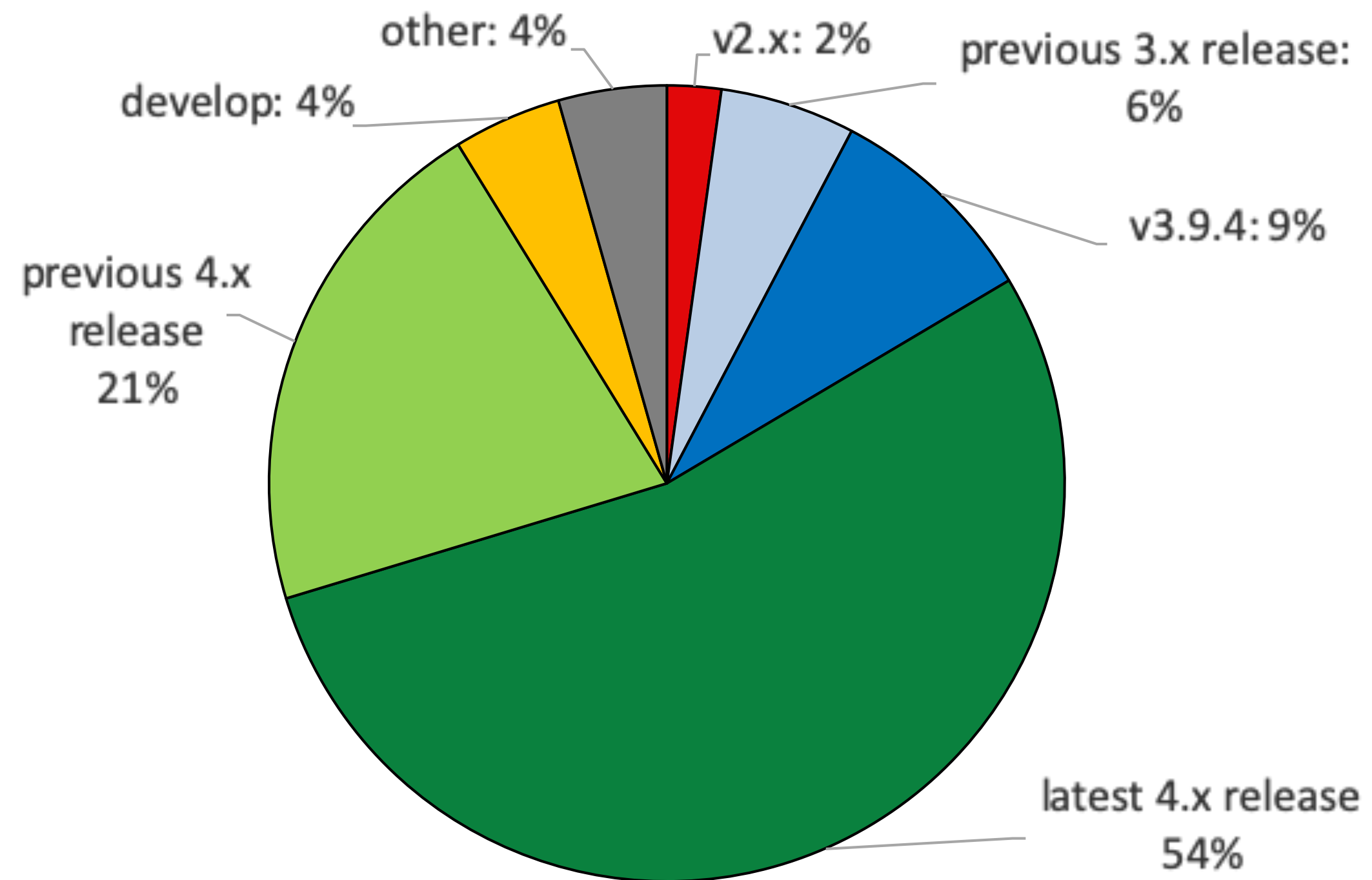


- large increase in "none" (18% ➡ 49%)
- less customising of easyblocks/easyconfigs
 - only easyconfigs: 37% ➡ 23%
 - both: 29% ➡ 12%
- customised framework remains stable at 16%



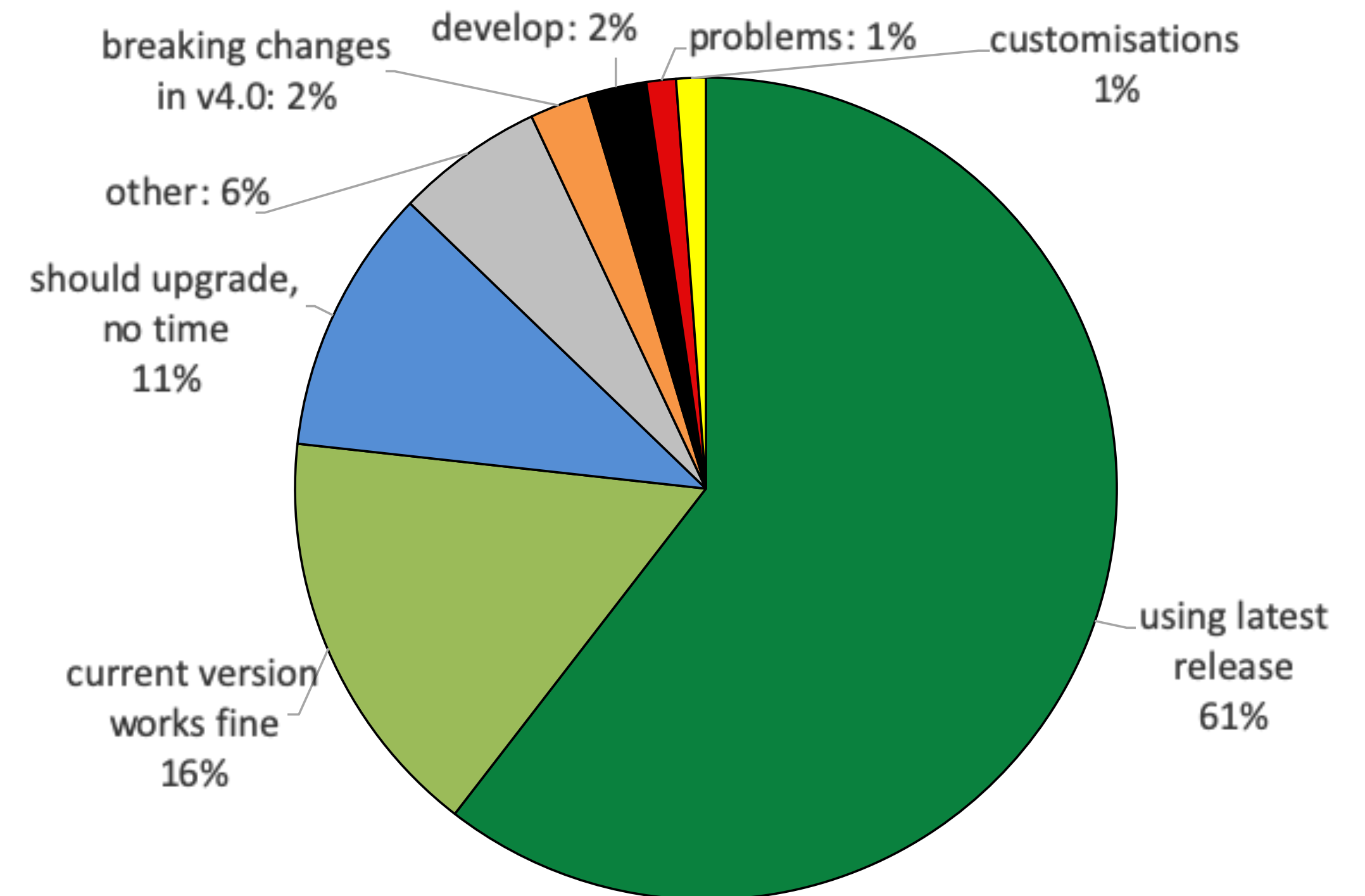
3rd easybuild User Survey

Which EasyBuild version do you use?



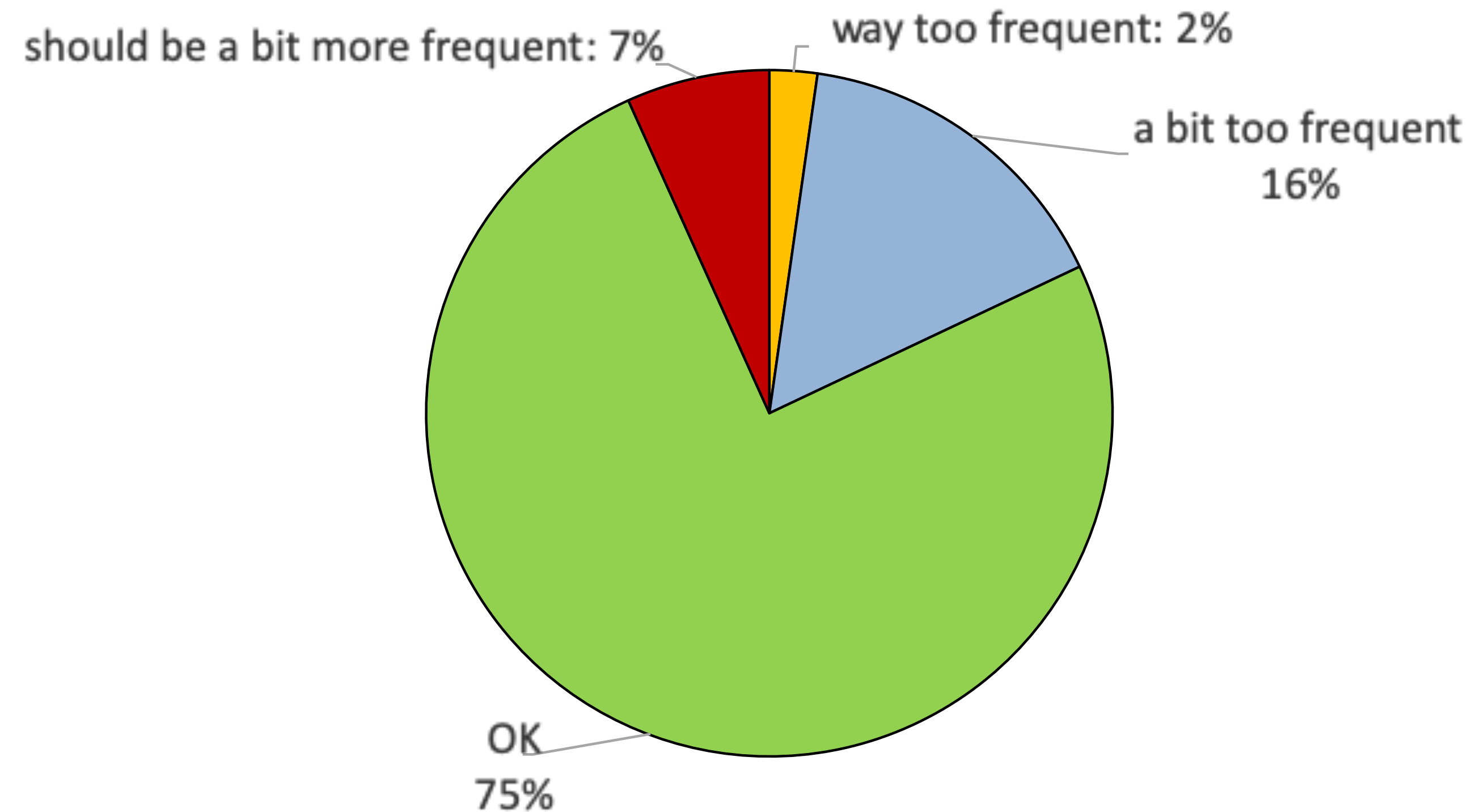
> 90% relies on EasyBuild releases,
75% is already using EasyBuild 4.x,
small minority using develop (only maintainers?)

If you're not using the latest release, why not?



overall very similar to previous survey
(small decrease in "problems with recent version")

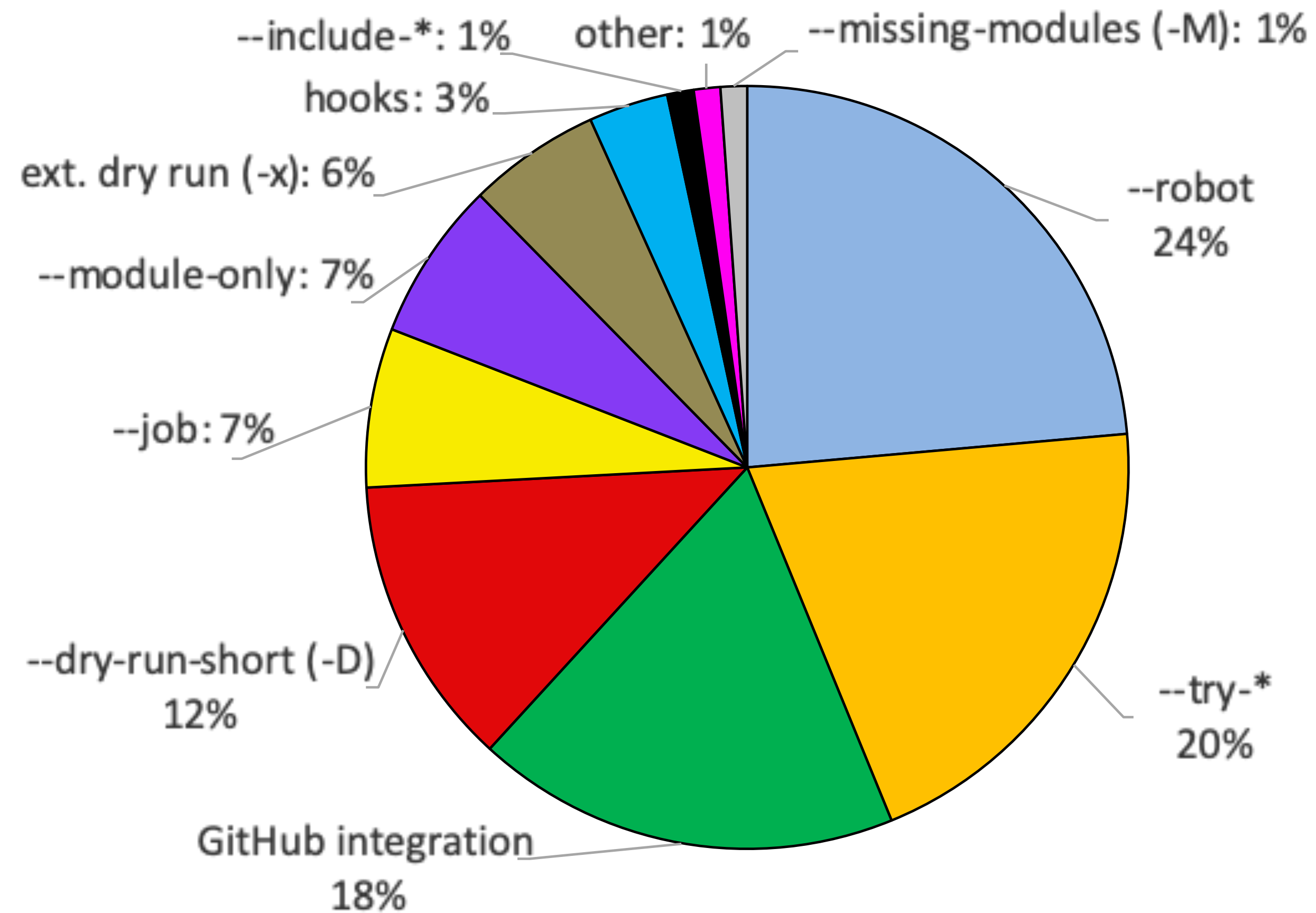
How do you like the frequency of EasyBuild releases?



vast majority like current release frequency,
slightly more "too frequent" than "bit more frequent"

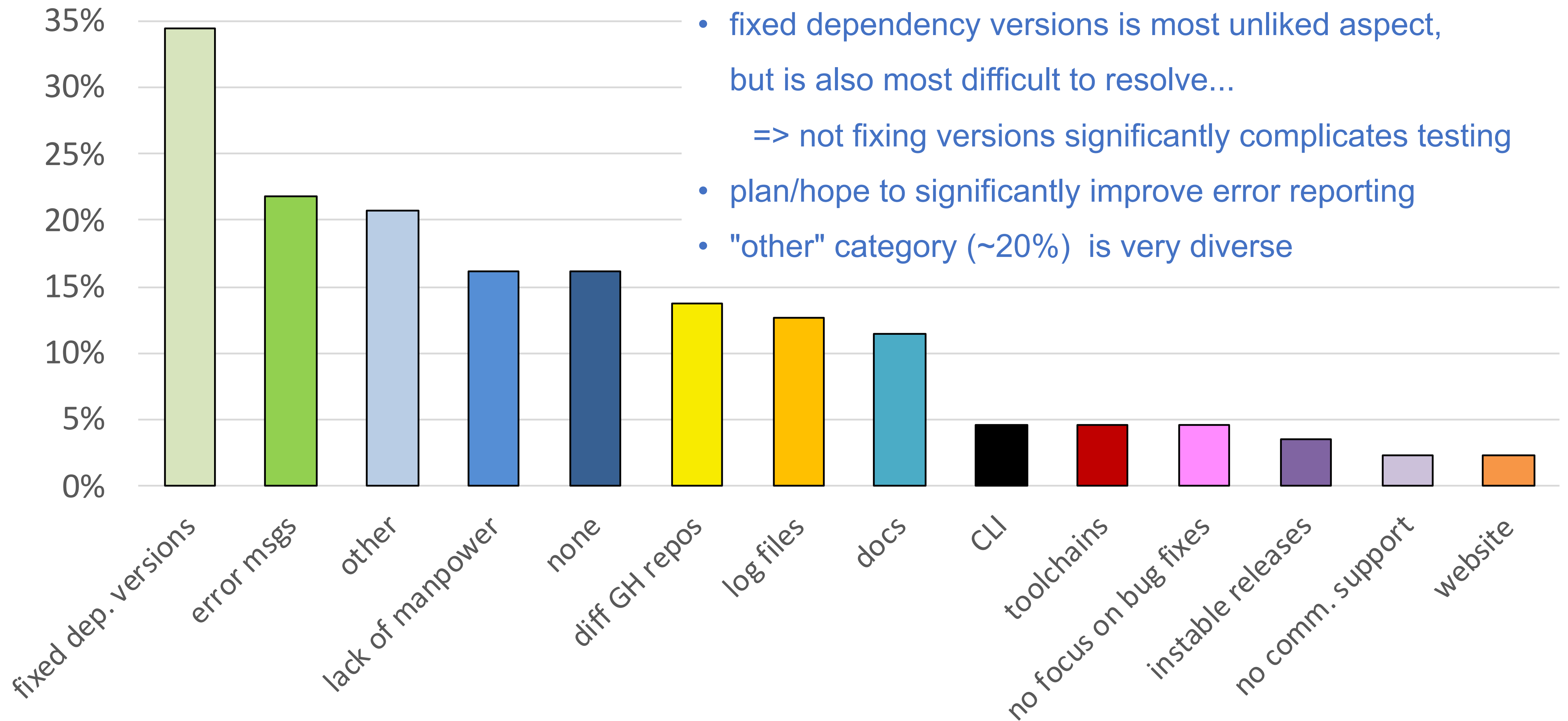
3rd easybuild User Survey

What is your favourite EasyBuild feature?



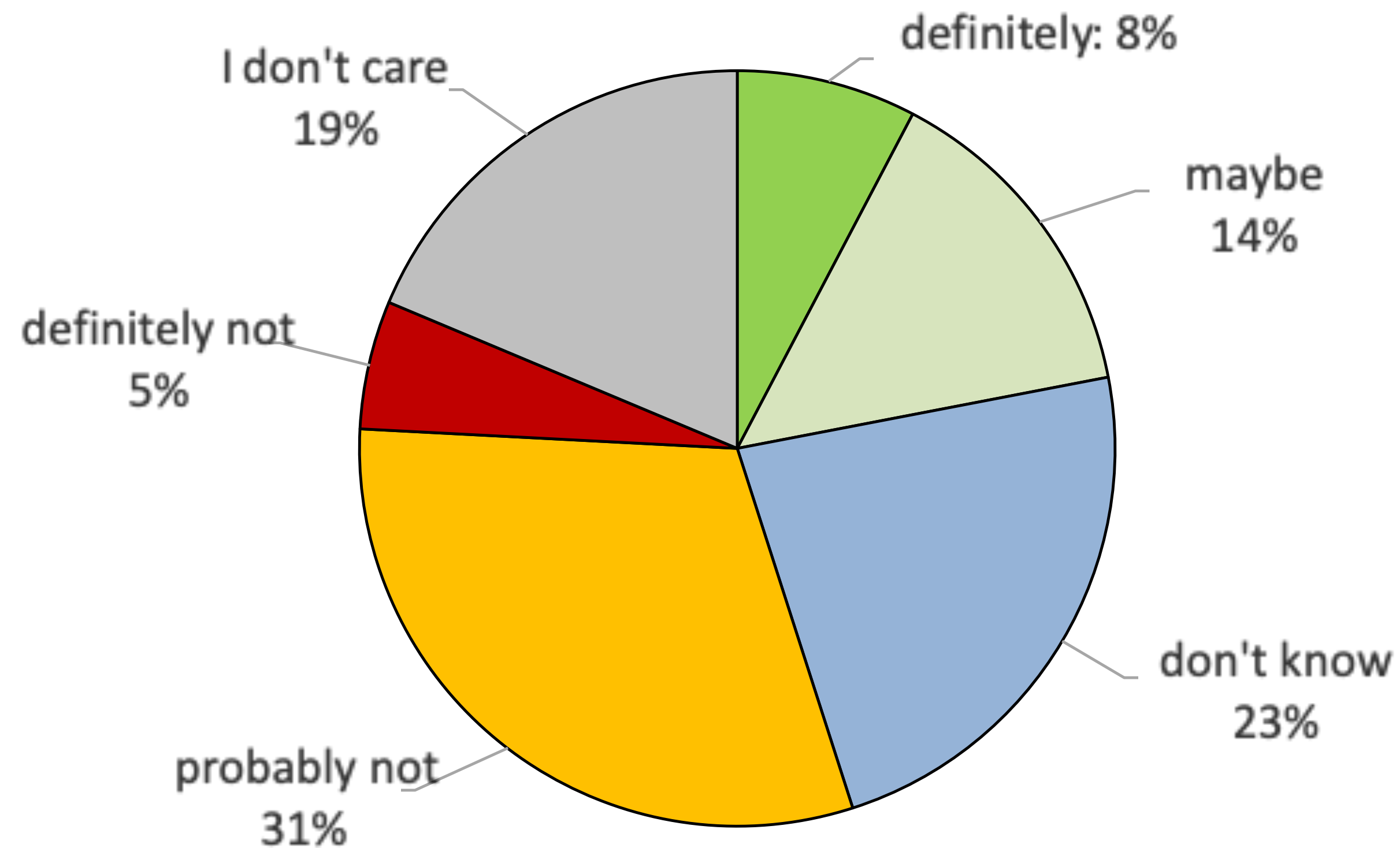
My personal favourite is definitely the GitHub integration (--new-pr & co).

Which parts of EasyBuild do you not like? (check all that apply)



3rd easybuild User Survey

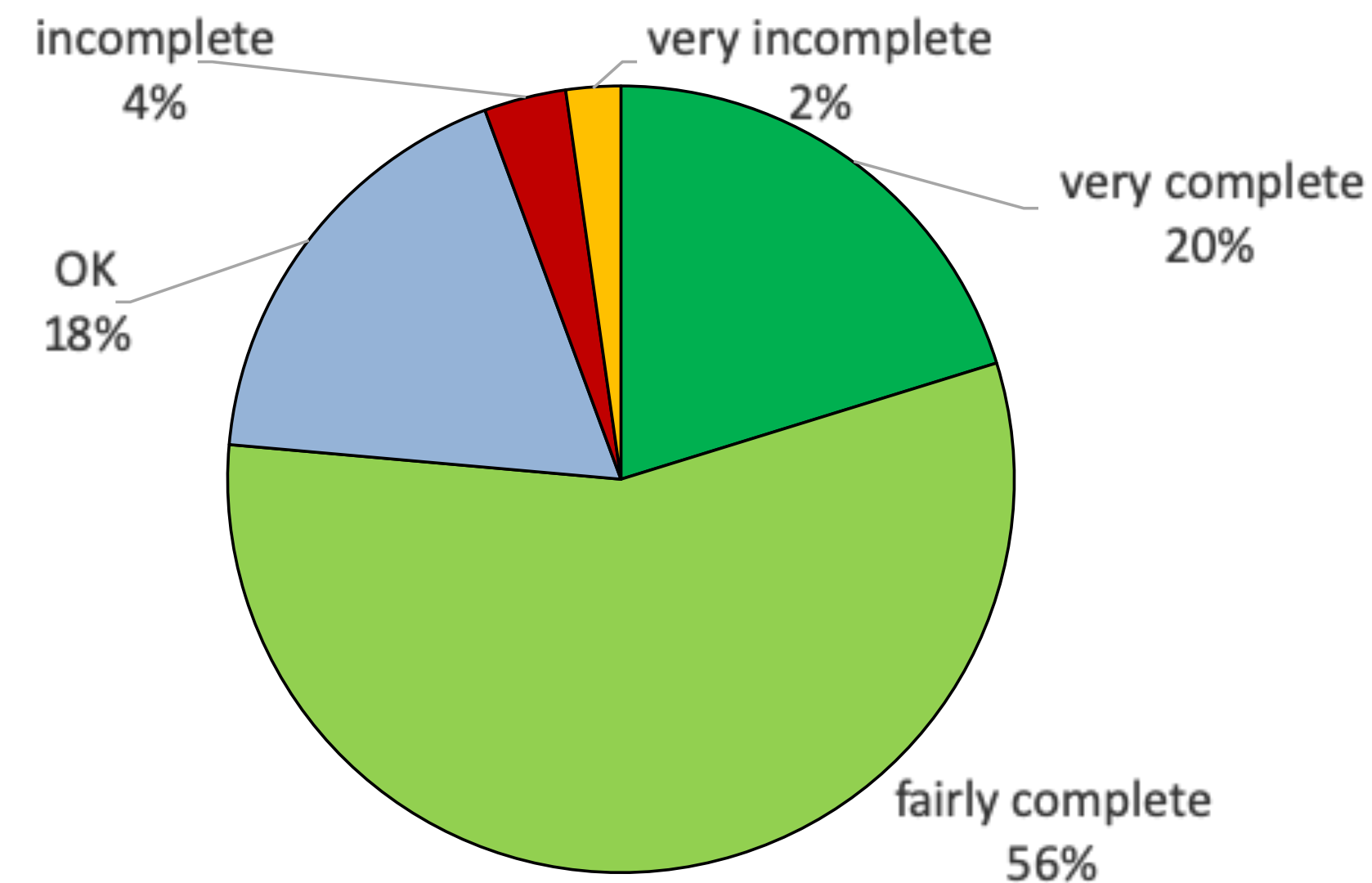
If a commercial support option would be available,
would you consider purchasing it?



slight increase in interest for commercial support

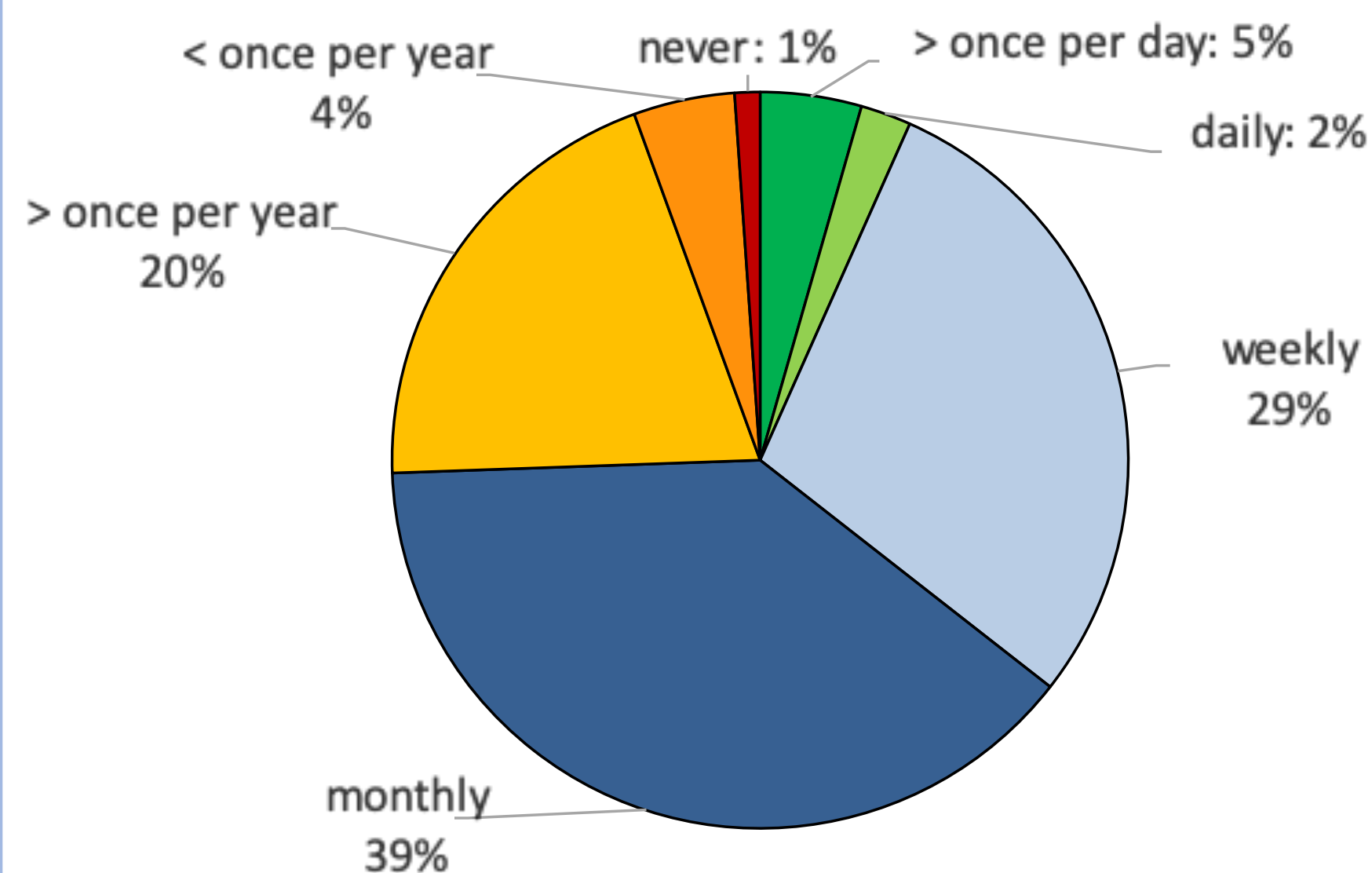
3rd easybuild User Survey

How complete is the EasyBuild documentation?



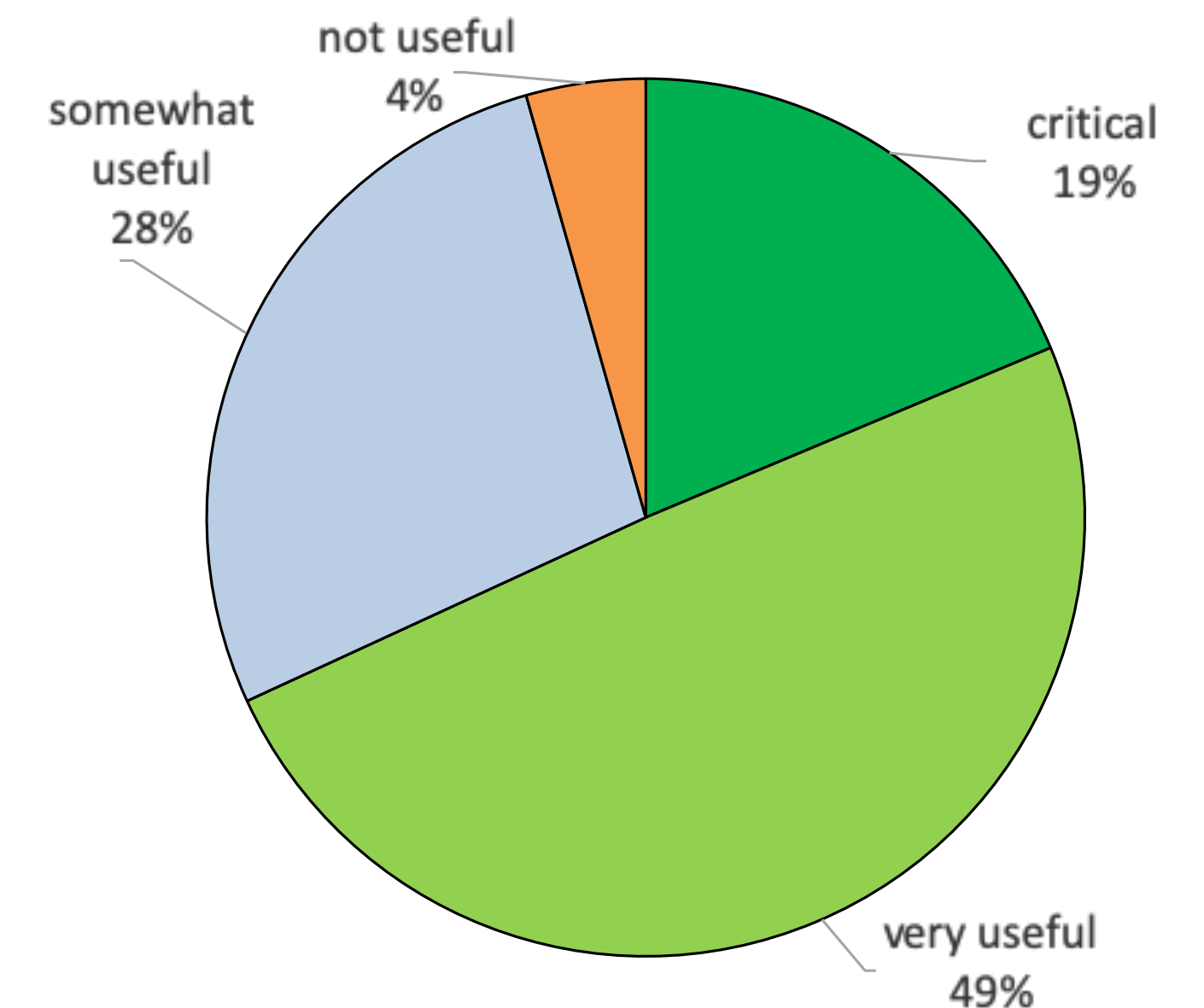
fairly to very complete
according to ~76% (stable)

How often do you consult the EasyBuild documentation?



~75% consults documentation
monthly or more frequent (stable)

How useful is the EasyBuild documentation to you?

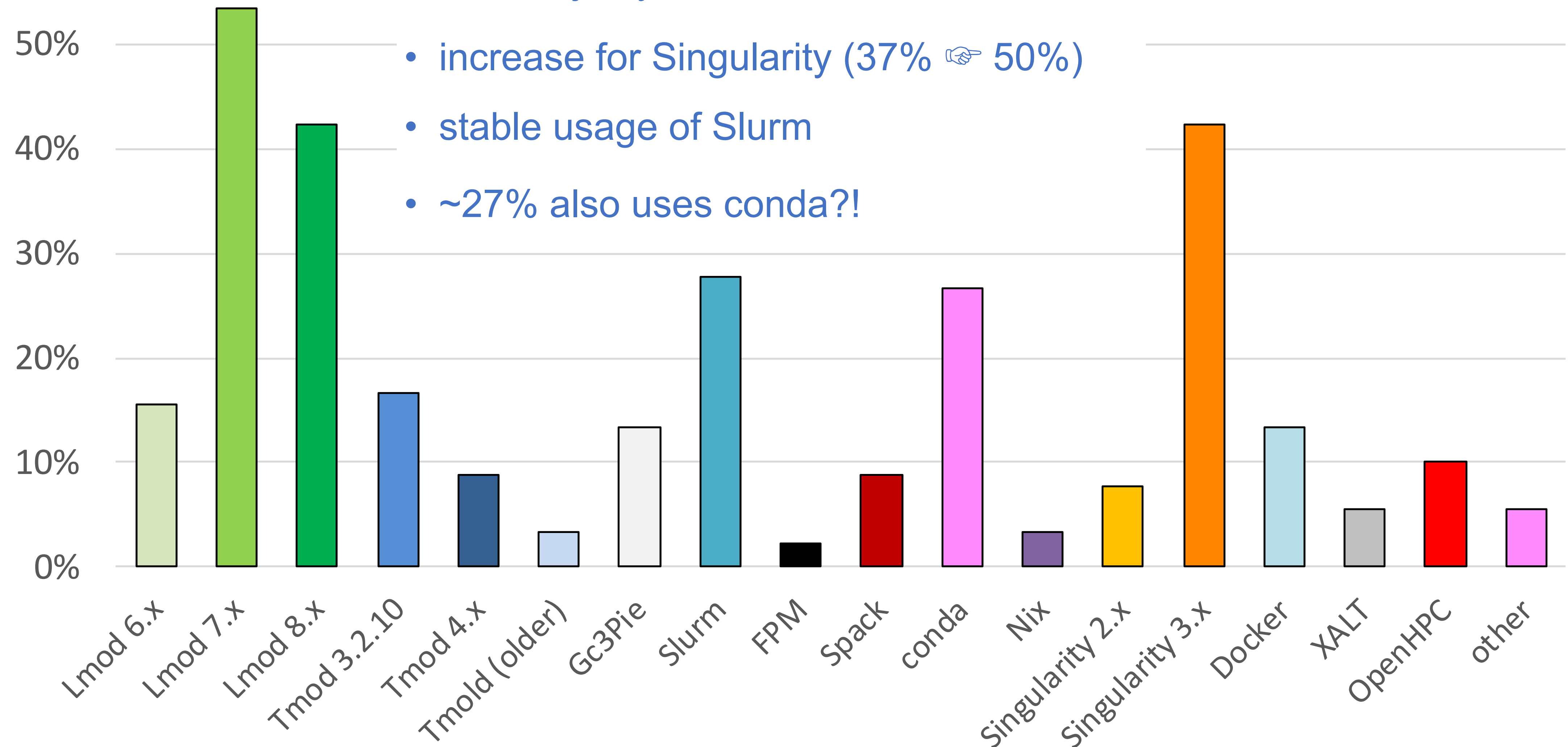


96% finds it at least
somewhat useful (stable)
(nobody answered "totally useless")

3rd easybuild User Survey

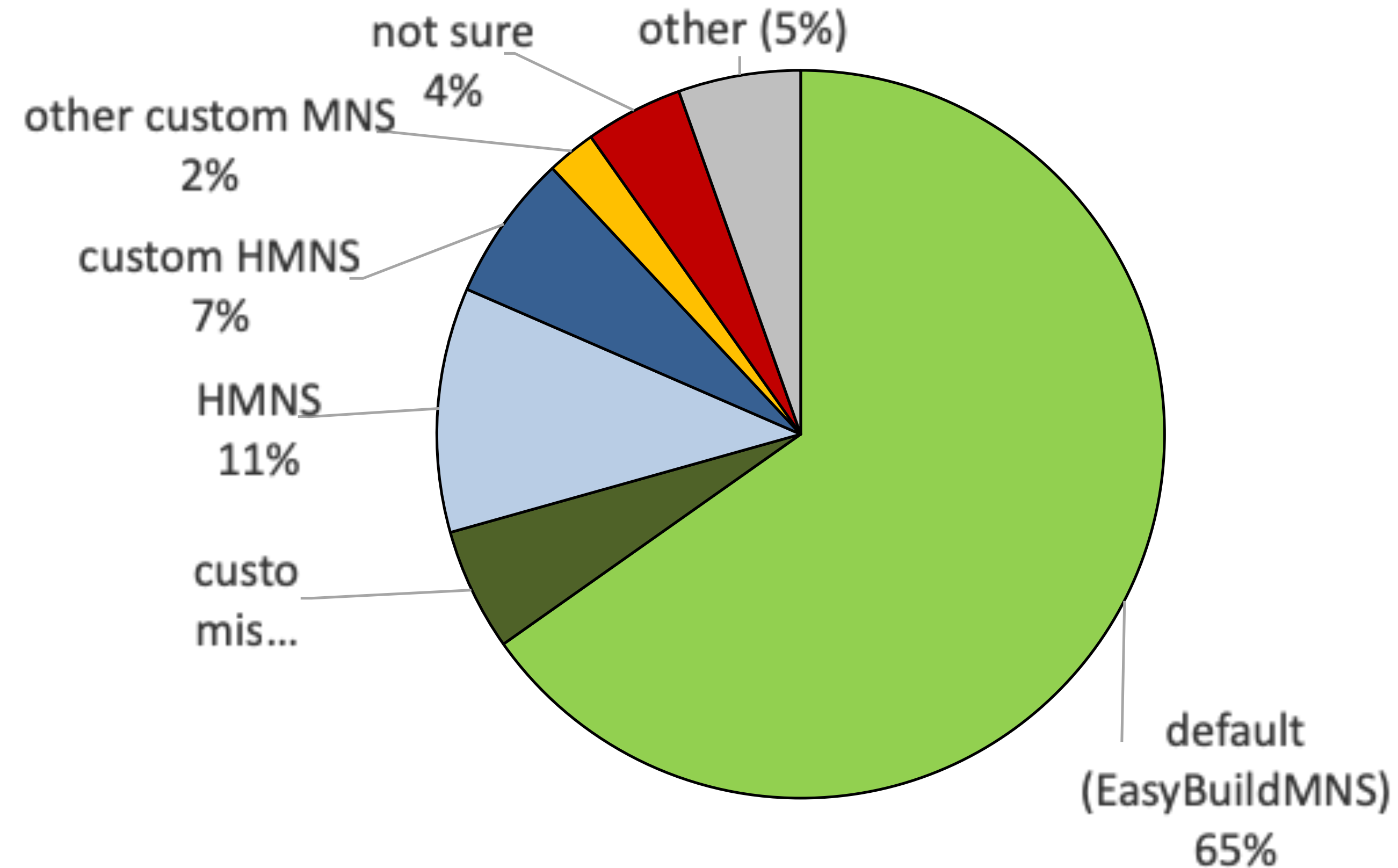
Do you use any other tools/projects in combination with EasyBuild?

- vast majority uses Lmod as modules tool
- increase for Singularity (37% ➡ 50%)
- stable usage of Slurm
- ~27% also uses conda?!



3rd easybuild User Survey

Which module naming scheme do you use?



minor decrease for EasyBuildMNS (65% ➡ 55%),

small shift towards custom MNS

Flat module naming scheme



legend

(not available)

(available)

(loaded)

- all modules are always available for loading
- long(er) module names
- 'module avail' may be overwhelming for users
- too easy to load incompatible modules together

GCC/5.3.0

GCC/6.1.0

OpenMPI/1.10.2-GCC-5.3.0

OpenMPI/2.1.0-GCC-5.3.0

OpenMPI/1.10.3-GCC-6.1.0

OpenMPI/2.1.0-GCC-6.1.0

FFTW/3.3.4-gompi-2016.04

FFTW/3.3.6-gompi-2016.04

FFTW/3.3.4-gompi-2016.07

FFTW/3.3.6-gompi-2016.07

Hierarchical module naming scheme (1)



- modules are organised in a tree-like fashion
- initially, only 'core' modules are available for loading
- typically 3 hierarchy levels: core, compiler-dependent, MPI-dependent
- other modules are only visible via 'module spider'

legend

(not available)

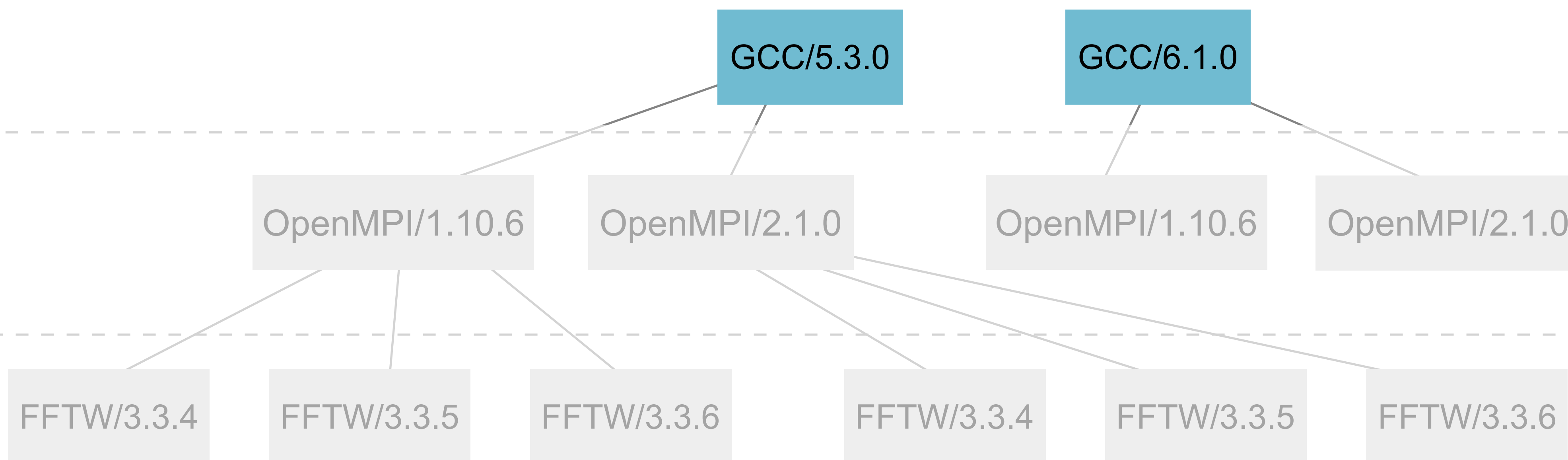
(available)

(loaded)

Core level

*Compiler level
(compiler-dependent)*

*MPI level
(MPI-dependent)*



Hierarchical module naming scheme (2)



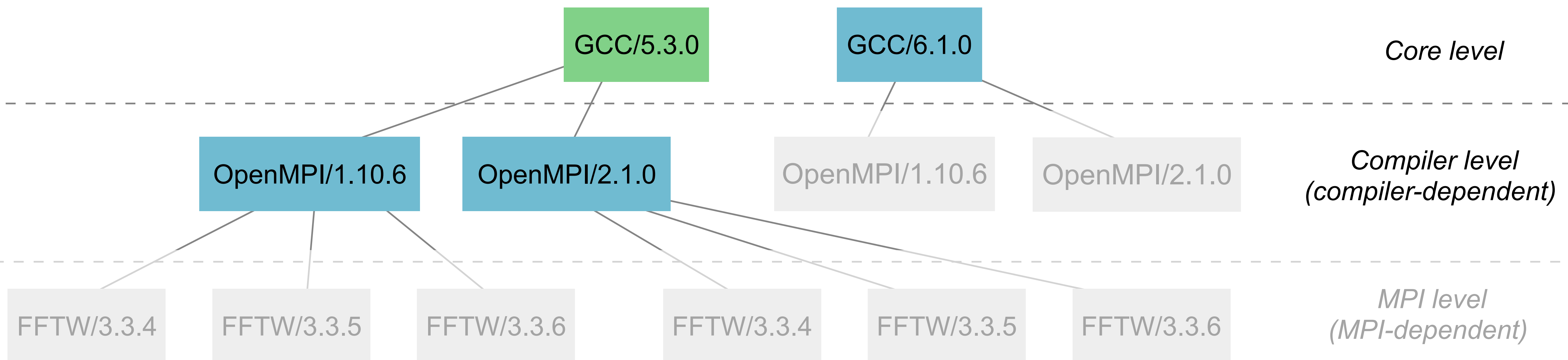
- Core modules may extend `$MODULEPATH` with an additional location
- loading a Core module may make more modules available
- in this example, loading a GCC module makes OpenMPI modules available

legend

(not available)

(available)

(loaded)



Hierarchical module naming scheme (3)



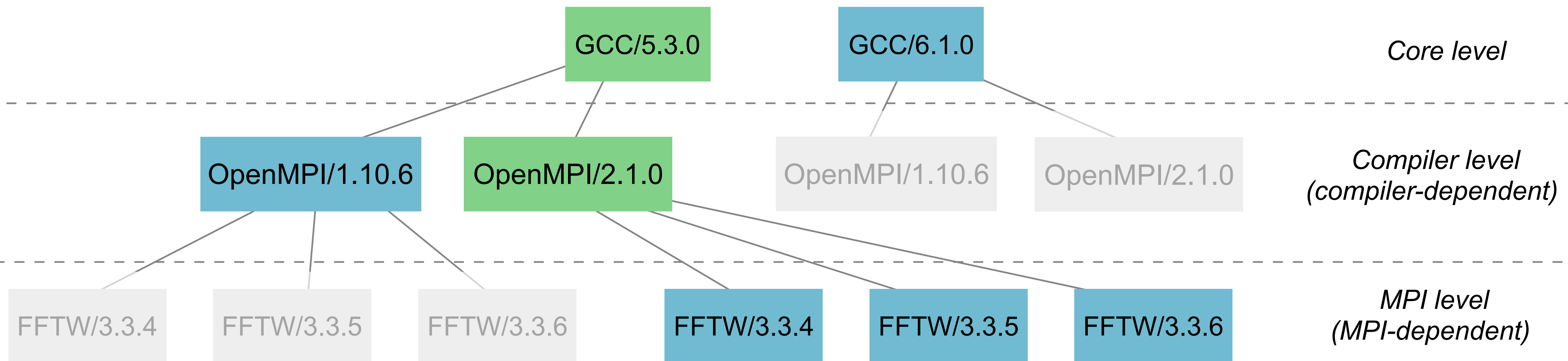
- even more modules may be made available by loading other modules
- for example, loading an OpenMPI modules reveals MPI-dependent modules
- EasyBuild can organise modules in hierarchy for you!

legend

(not available)

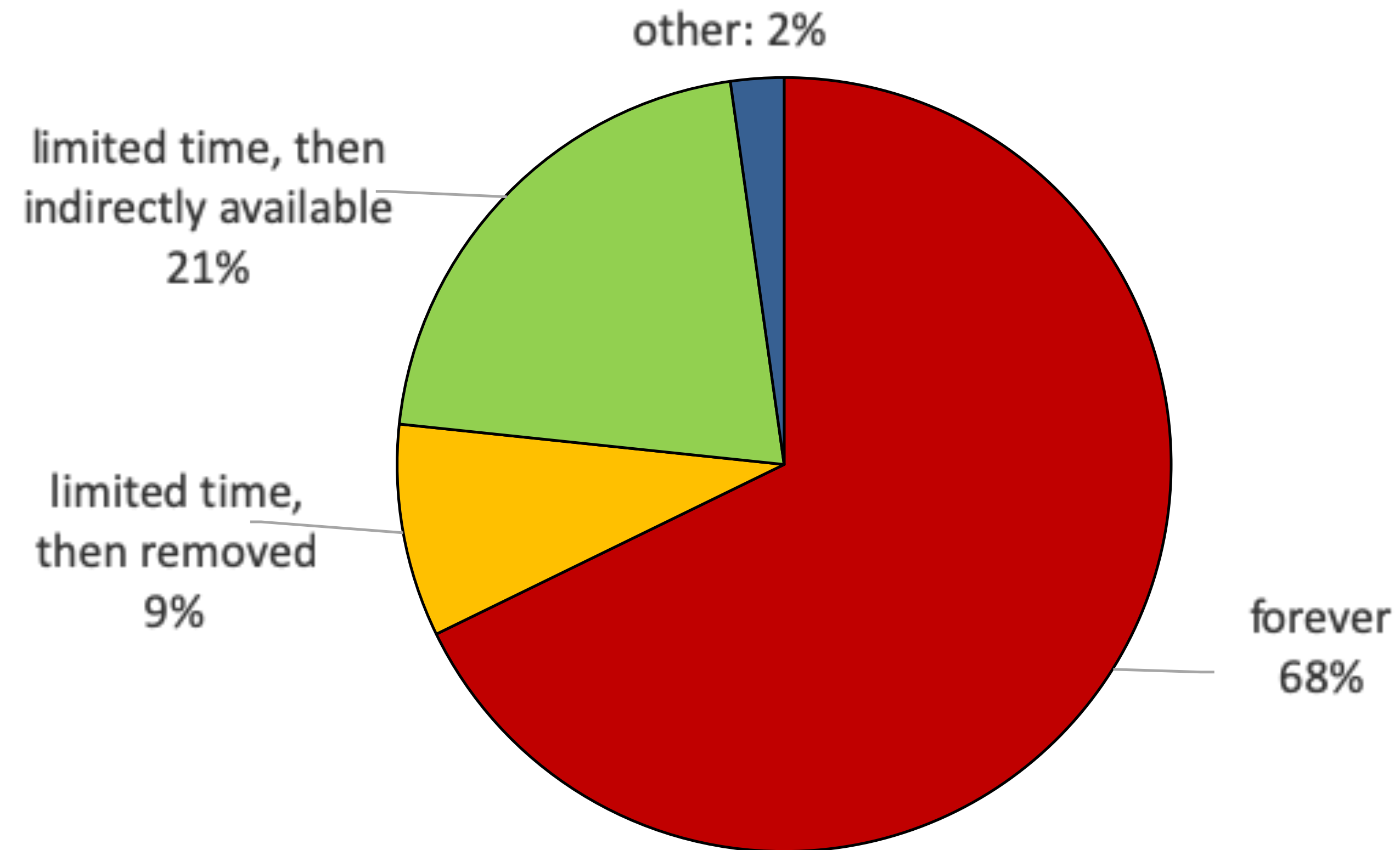
(available)

(loaded)



3rd easybuild User Survey

How long are modules installed with EasyBuild available to users?

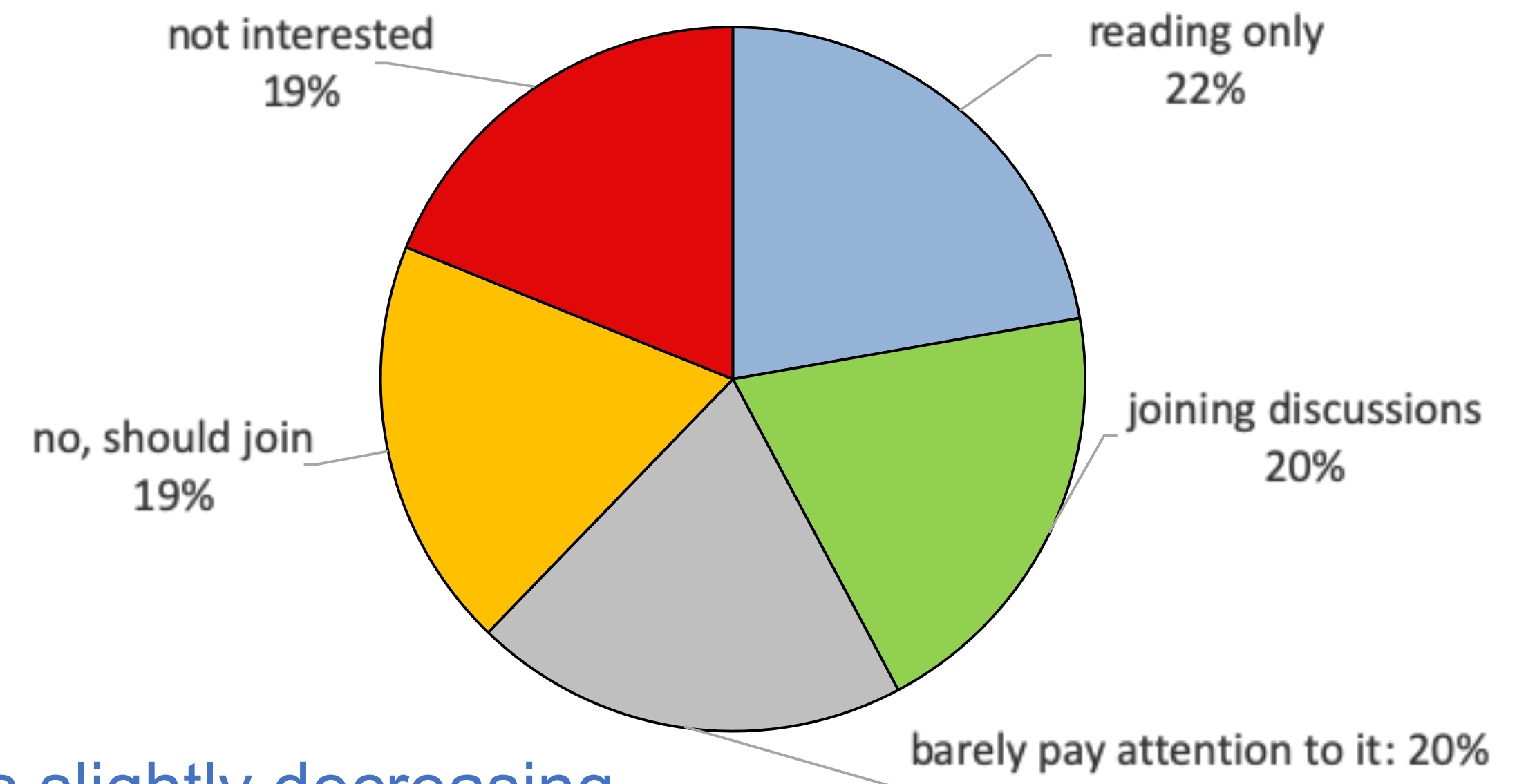
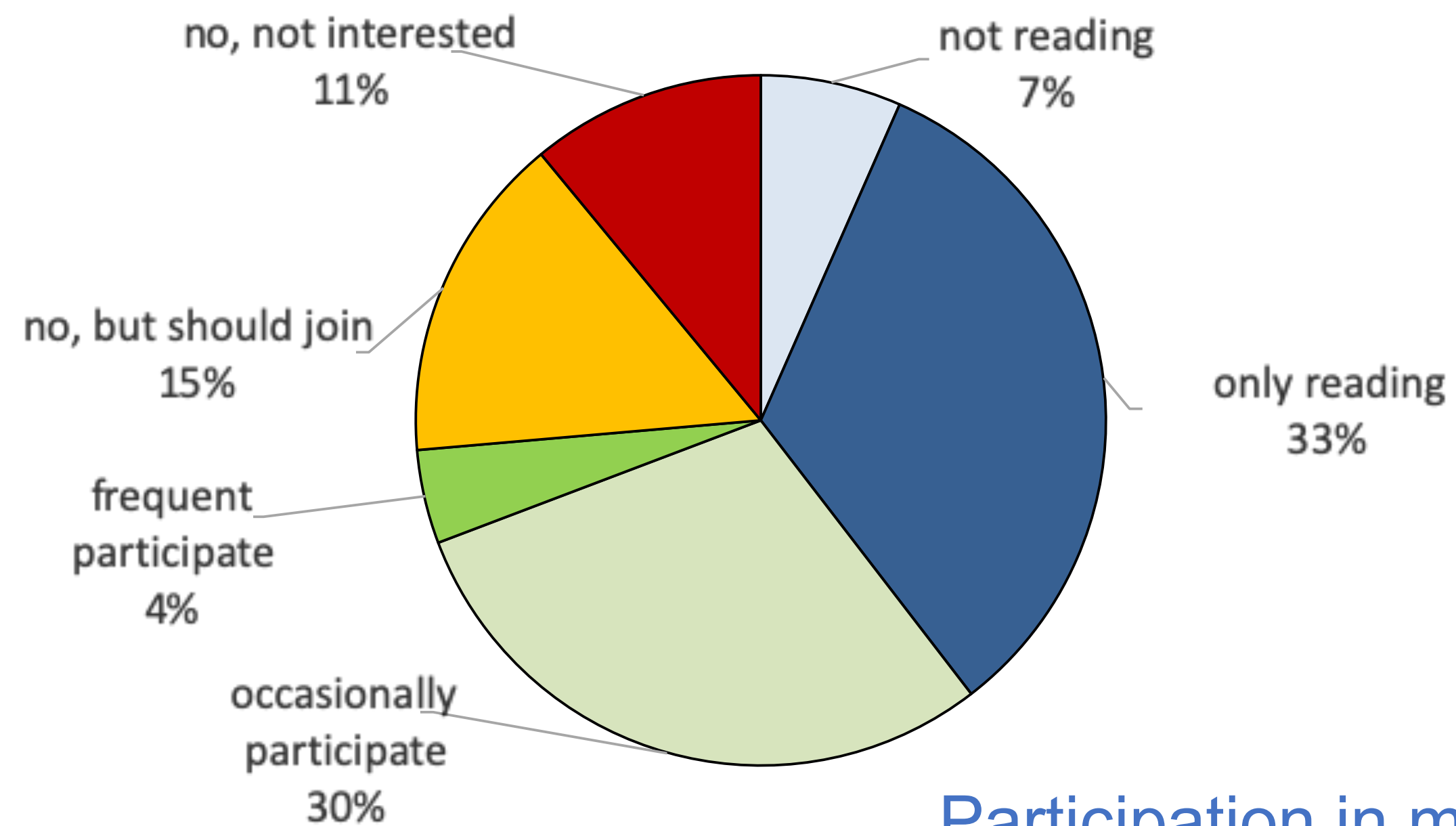
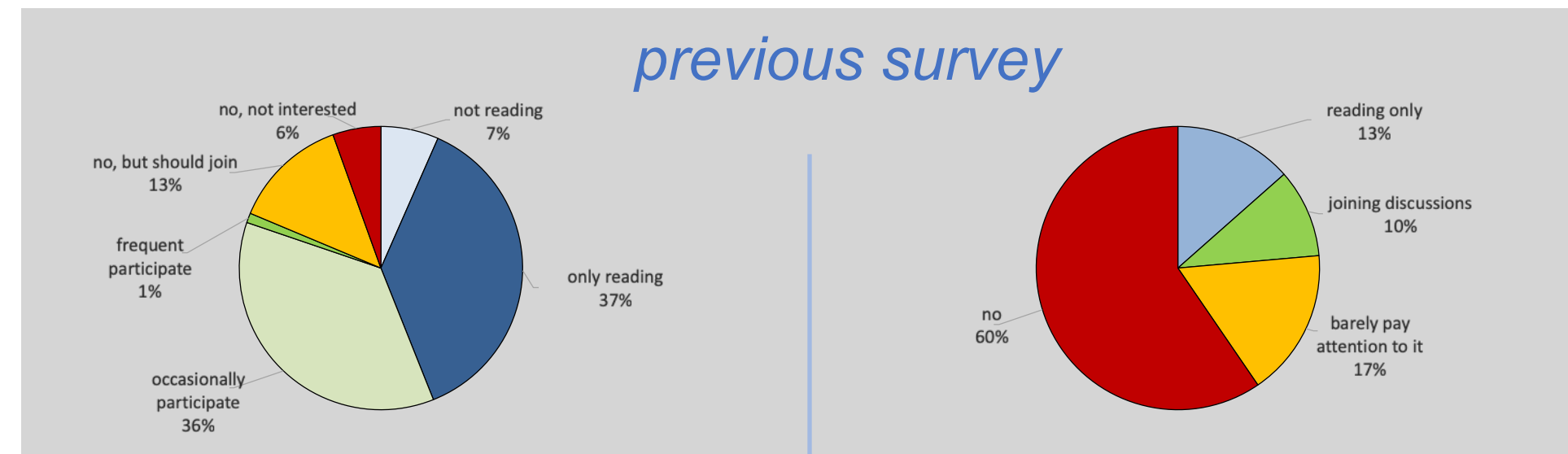


(almost identical to results of previous survey)

3rd easybuild User Survey

Are you subscribed to the EasyBuild mailing list?

Do you use the EasyBuild IRC/Slack channel?

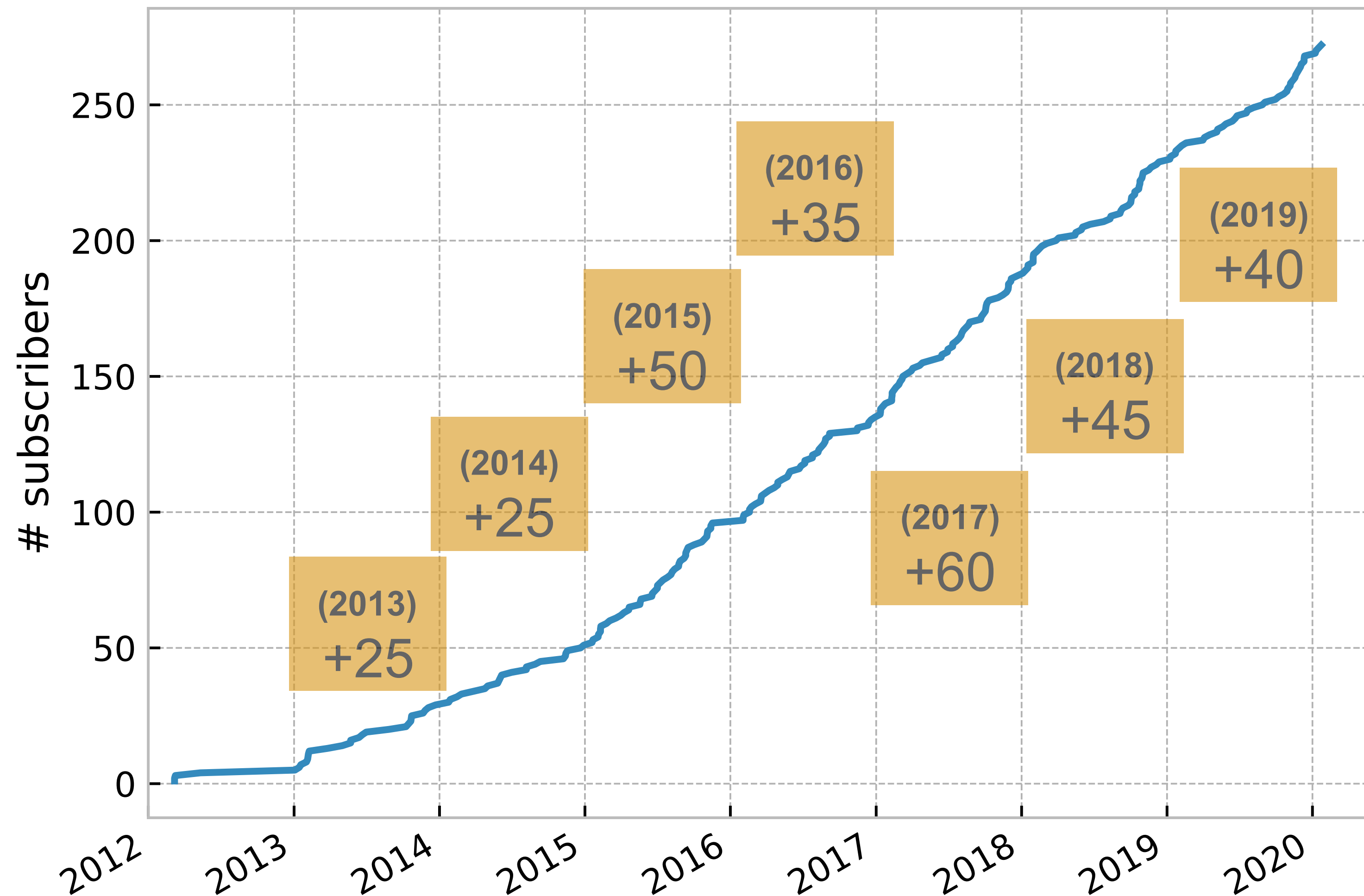


Participation in mailing list is slightly decreasing,
interest/use of Slack is significantly increasing.

A significant part of the EasyBuild community remains silent.

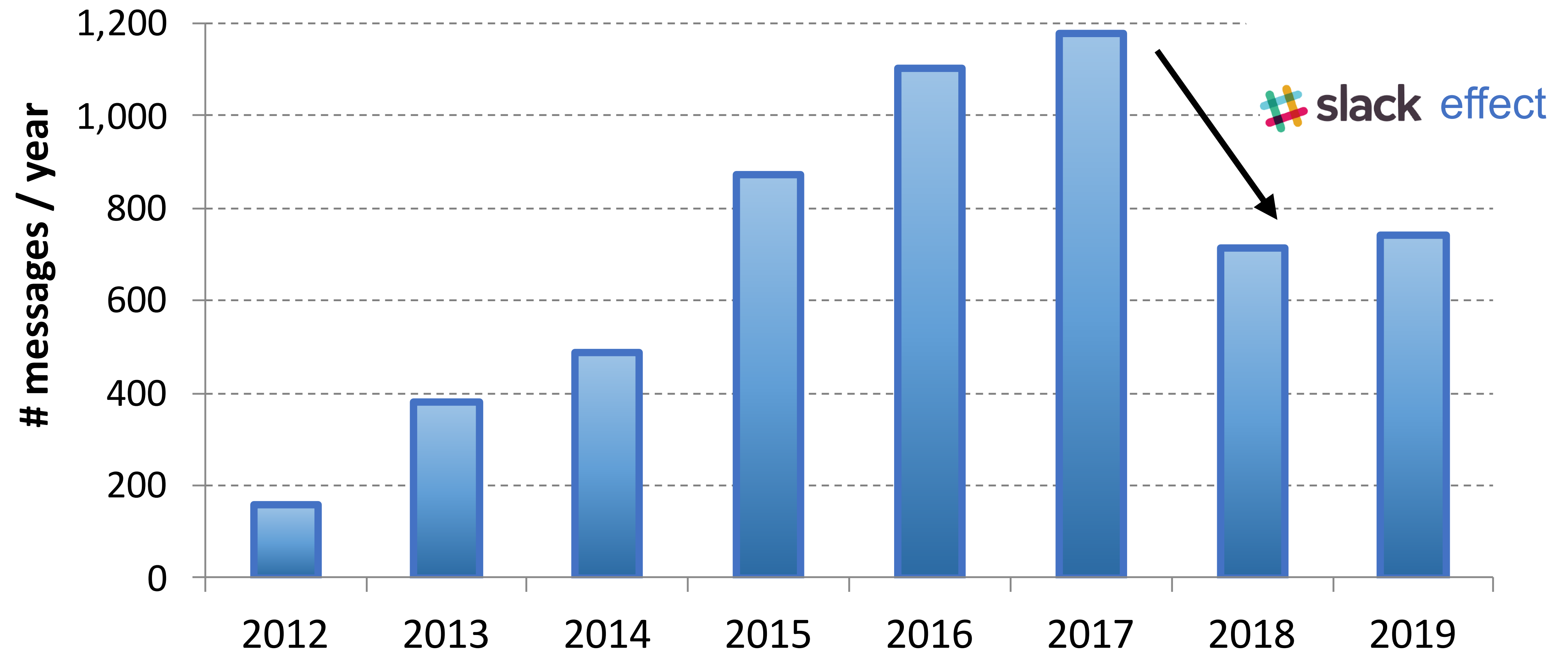
easybuild mailing list (subscribers)

Over 270 people subscribed to EasyBuild mailing list, still growing...



easybuild mailing list (traffic)

Traffic on EasyBuild mailing list peaked in 2017,
due to start of EasyBuild Slack channel, but remains fairly active.

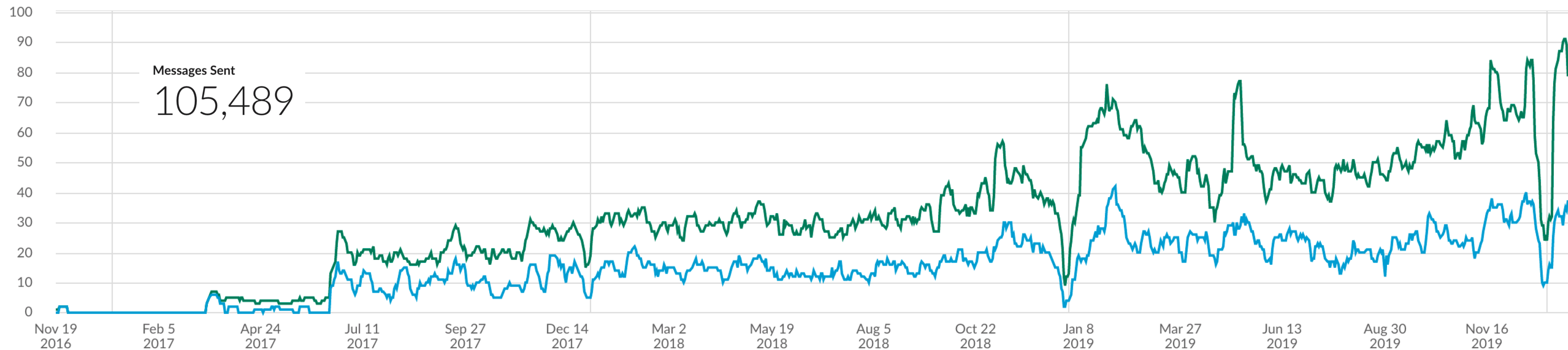


easybuild slack channel (+ IRC)

- > 200 members on Slack channel (+70 in 2019!)
- steady growth in activity since start of EasyBuild Slack in 2017

self-invite via <https://easybuild-slack.herokuapp.com> !

Weekly Daily



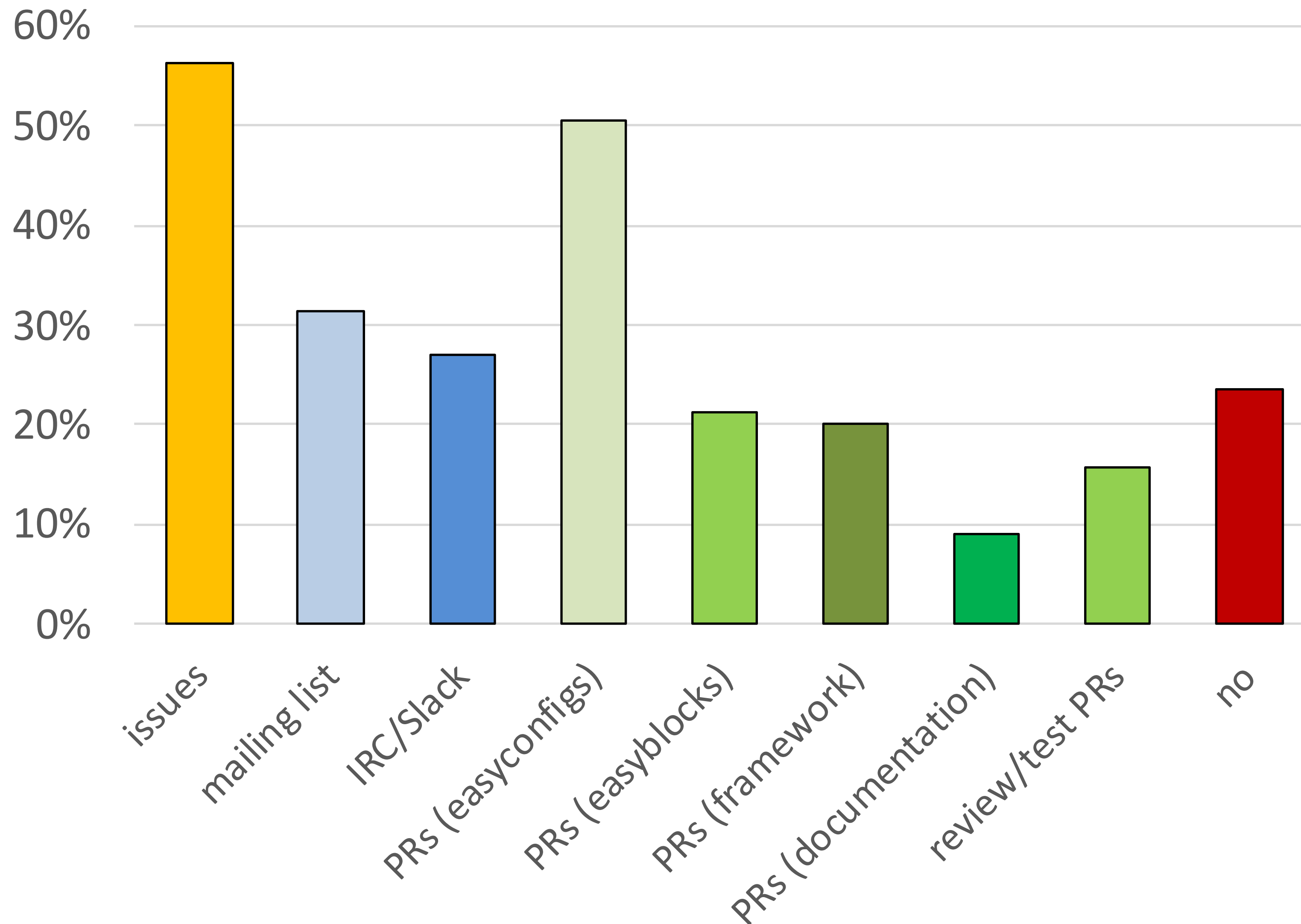
● Weekly active members ● Members who posted

Contributors, contributors, contributors...



3rd easybuild User Survey

Do you actively contribute to EasyBuild? (check all that apply)

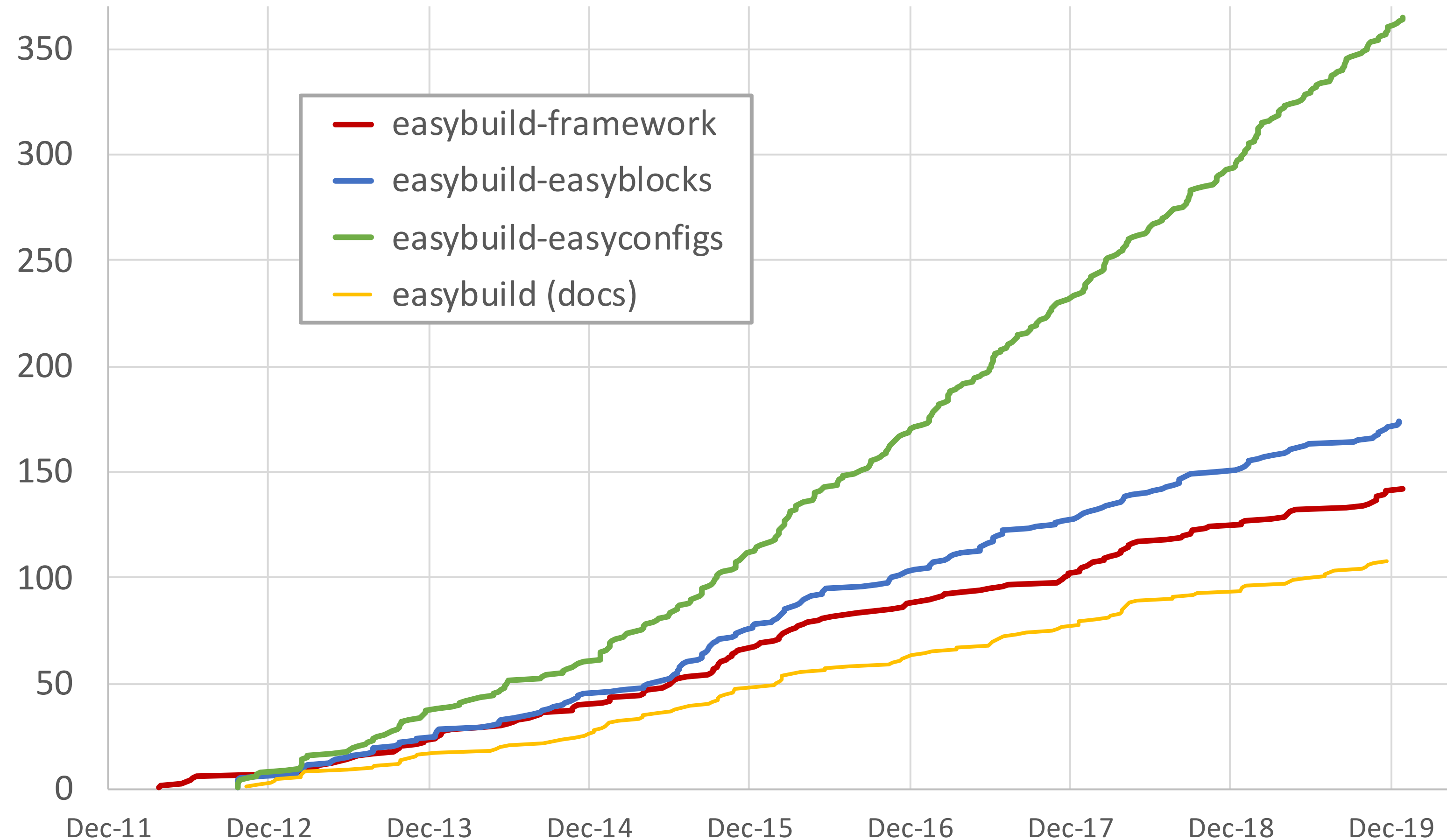


- issues (53% ➡ 56%)
- mailing list (23% ➡ 31%)
- **IRC/Slack (13% ➡ 27%)**
- **easyconfig PRs (38% ➡ 51%)**
- easyblock PRs (18% ➡ 21%)
- framework PRs (13% ➡ 20%)
- documentation PRs (9% ➡ 9%)
- review/testing of PRs (11% ➡ 16%)
- **no contributions (34% ➡ 24%)**

Overview of forks on GitHub

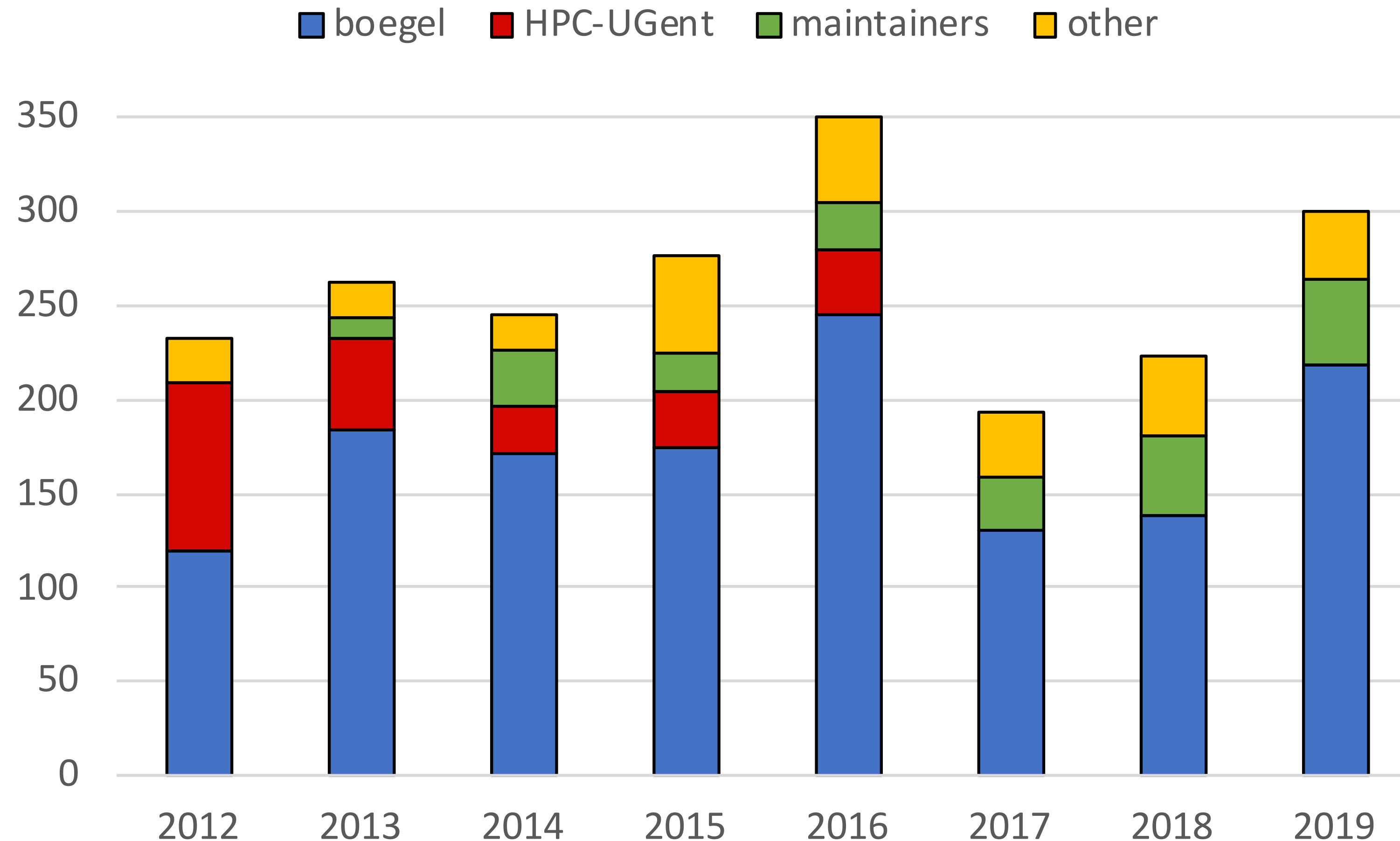


of (direct) forks of EasyBuild GitHub repositories



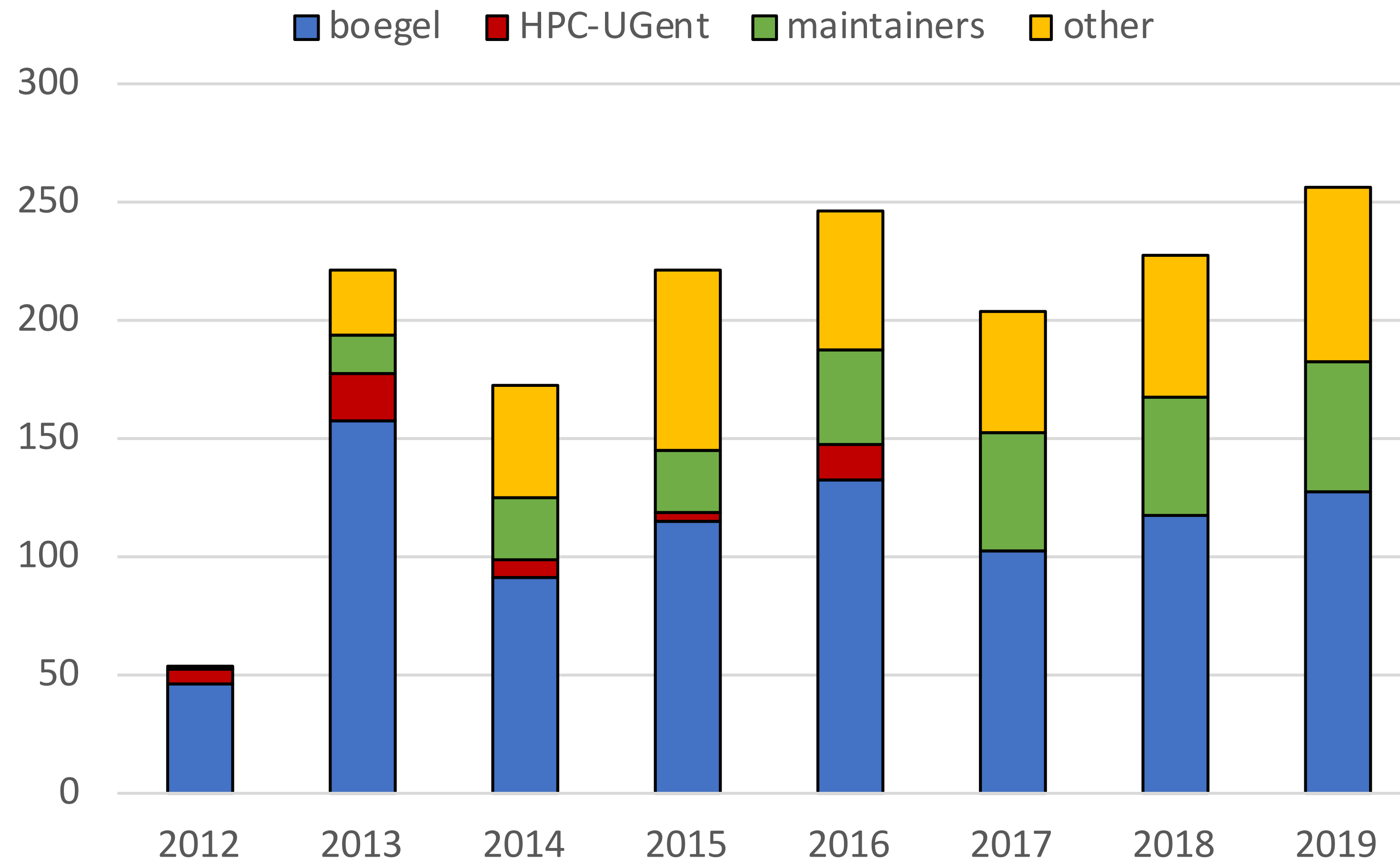
- actual fork count is higher (incl. indirect forks): ~415 for easyconfigs
- steady growth rate over the years across all repositories

Overview of contributions (framework)



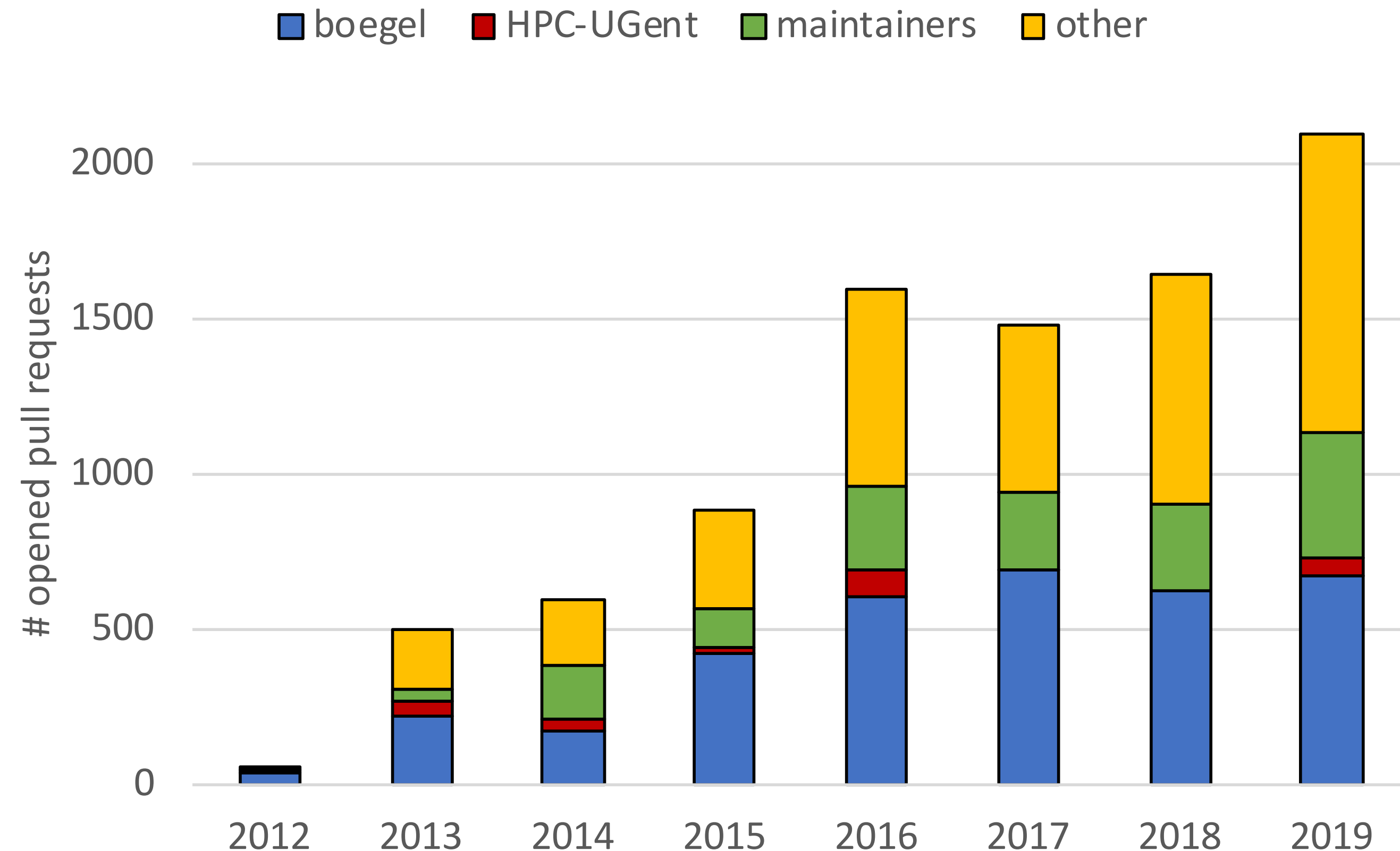
- effort of porting to Python 3 is reflected in PR count for 2019
- ~27% of PRs by contributors outside HPC-UGent

Overview of contributions (easyblocks)



- 21% of PRs by maintainers other than boegel (50%) + additional 29% by others
- small increase in PRs by contributors outside of maintainers (59 ➡ 73)

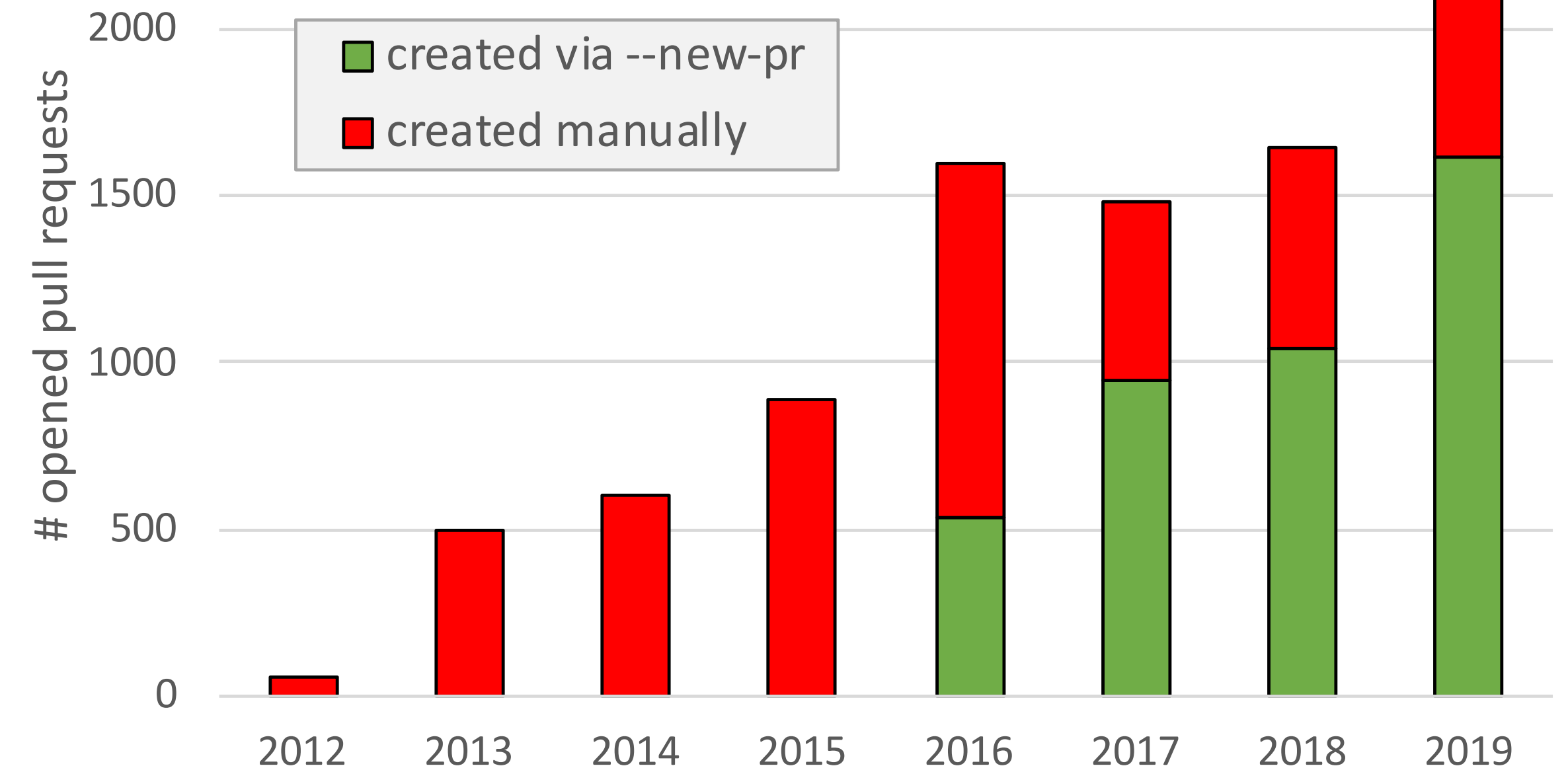
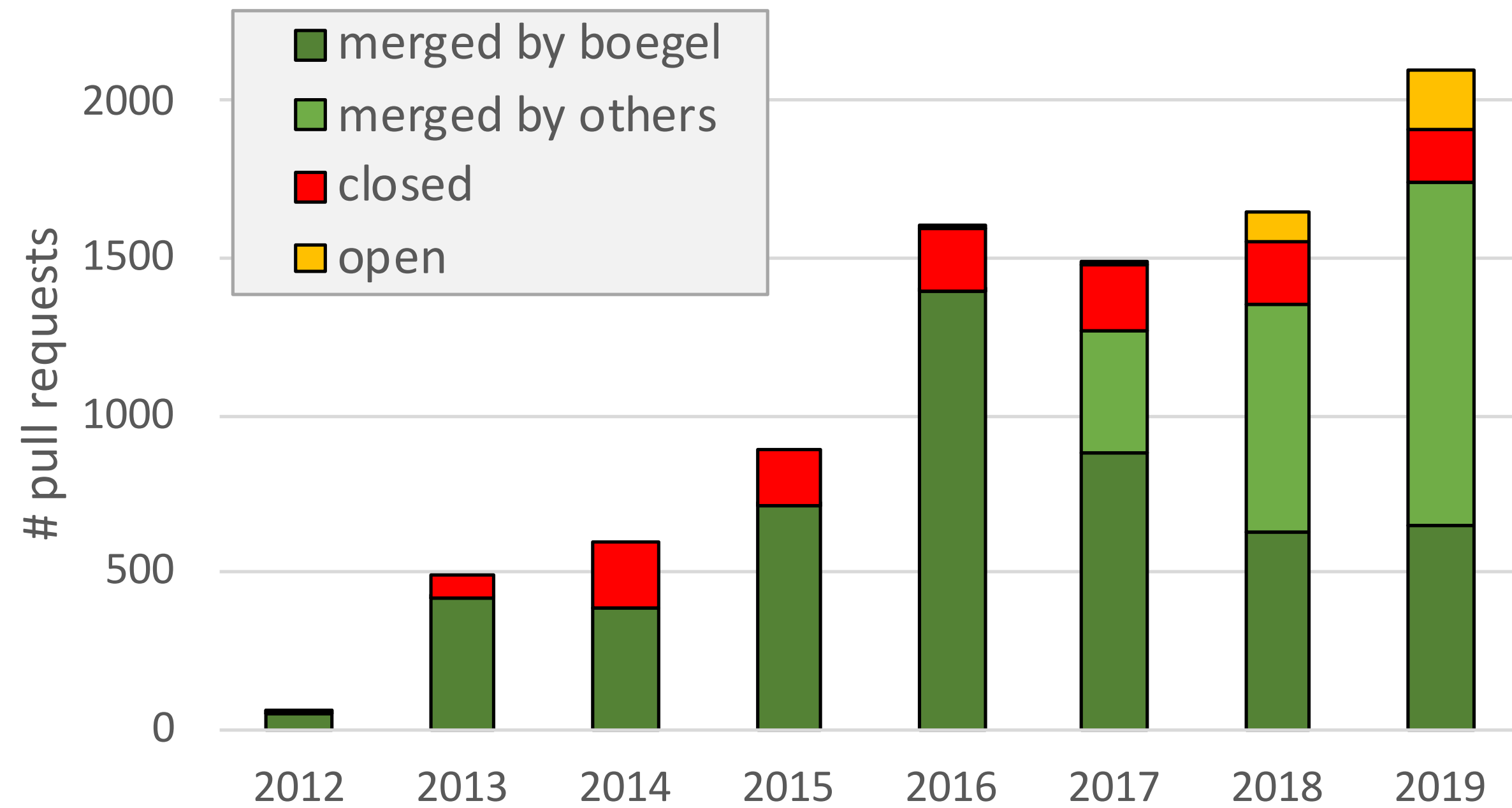
Overview of contributions (easyconfigs)



This is nuts.

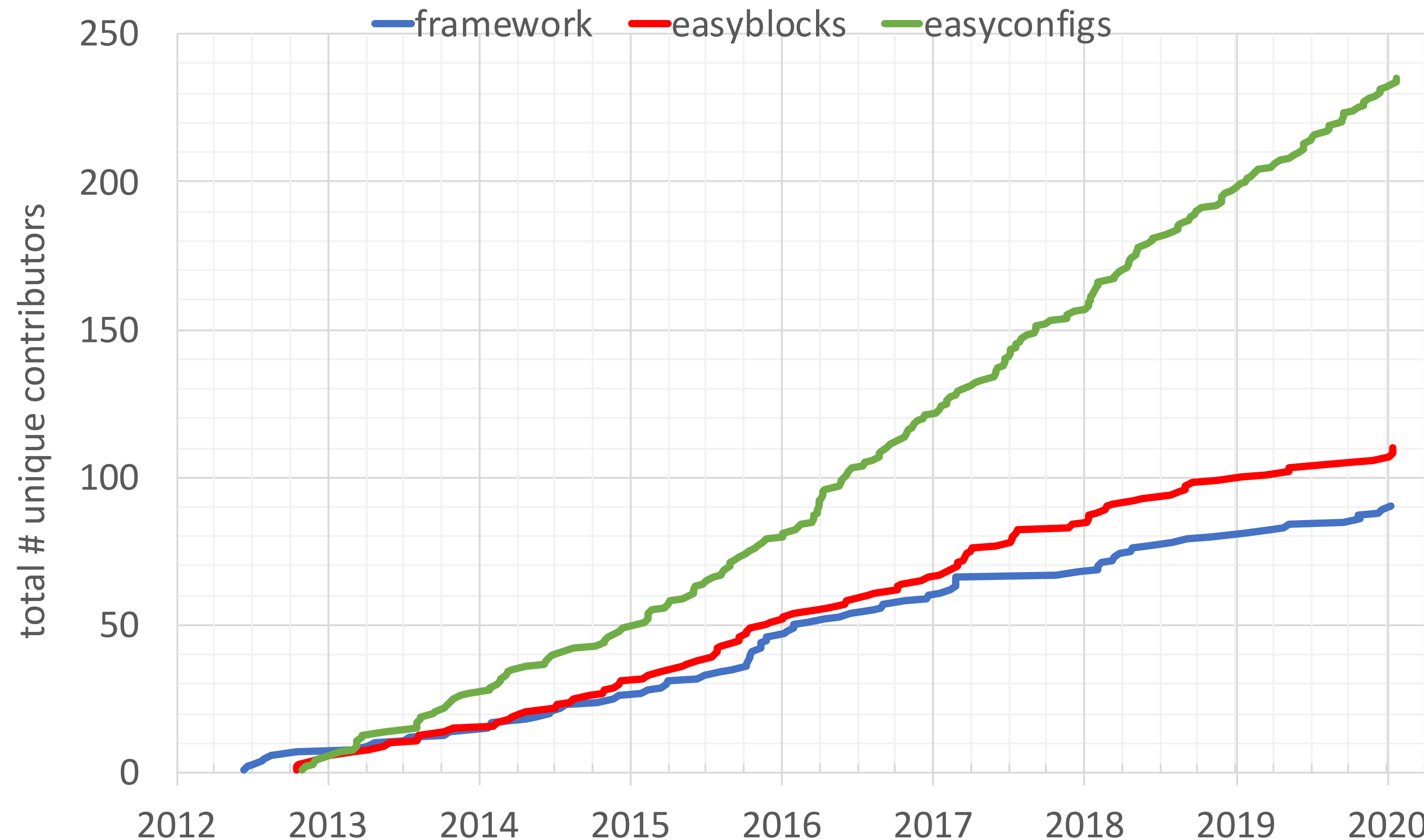
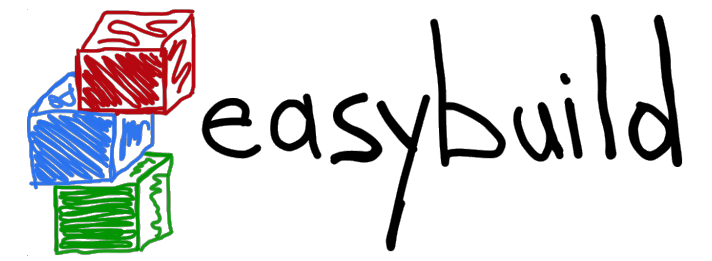
- explosive growth in easyconfig PRs in 2016 (partially thanks to `--new-pr`)
- stabilised in 2016-2018, significant increase in 2019: > 2,000 PRs...
- growing ratio of PRs from contributors that are not maintainers (~46% in 2018)

Overview of contributions (easyconfigs)



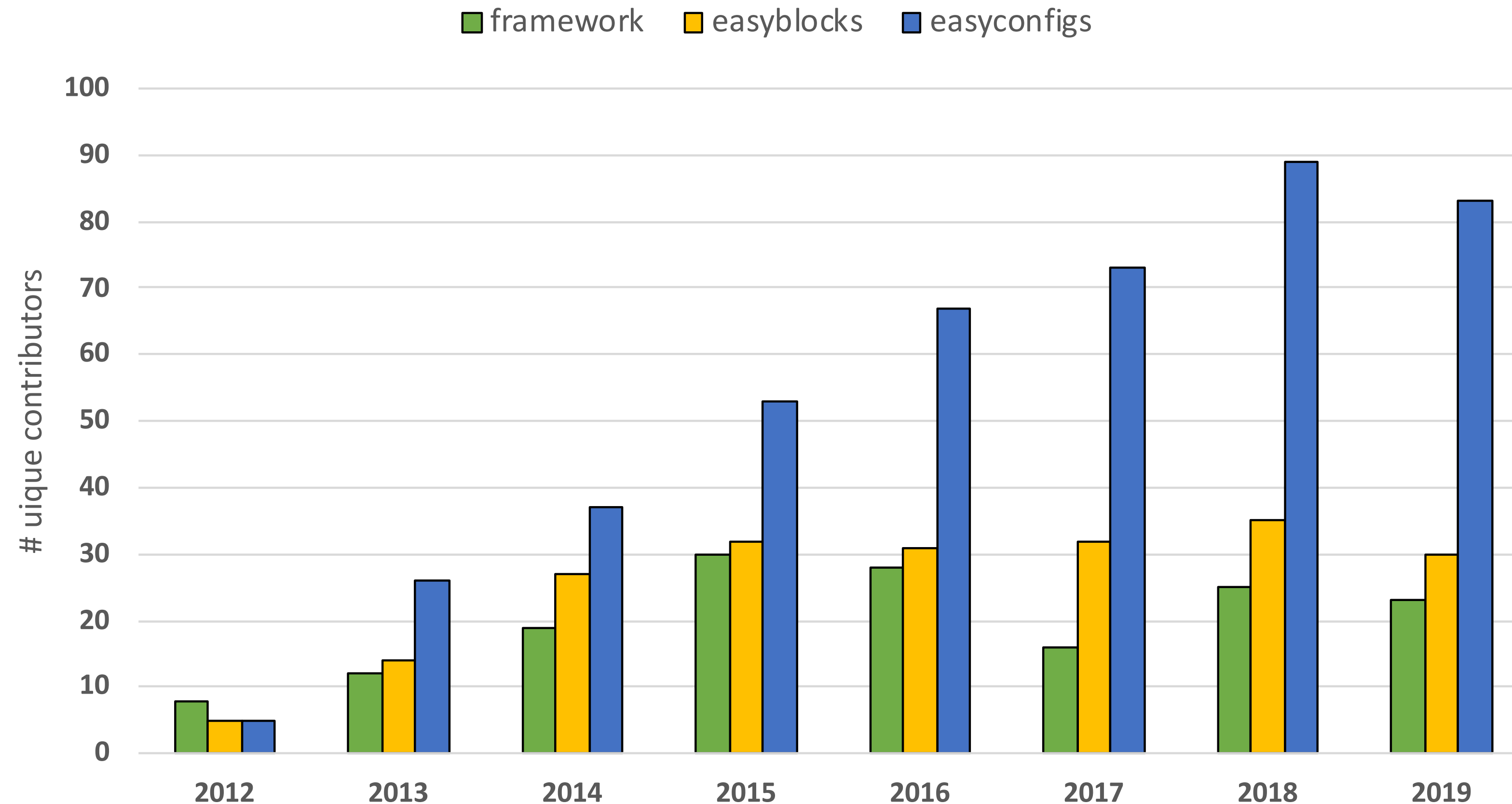
- growing number of easyconfig PRs tackled by other maintainers (~52% in 2019)
- ratio of PRs opened with "eb --new-pr" is increasing (77% in 2019)

Overview of unique contributors (total)



- 90 unique contributors to EasyBuild framework so far (+10 in 2019)
- continued growth in unique contributors for easyblocks (110)
- steadily growing number of unique contributors to easyconfigs (soon > 250?)

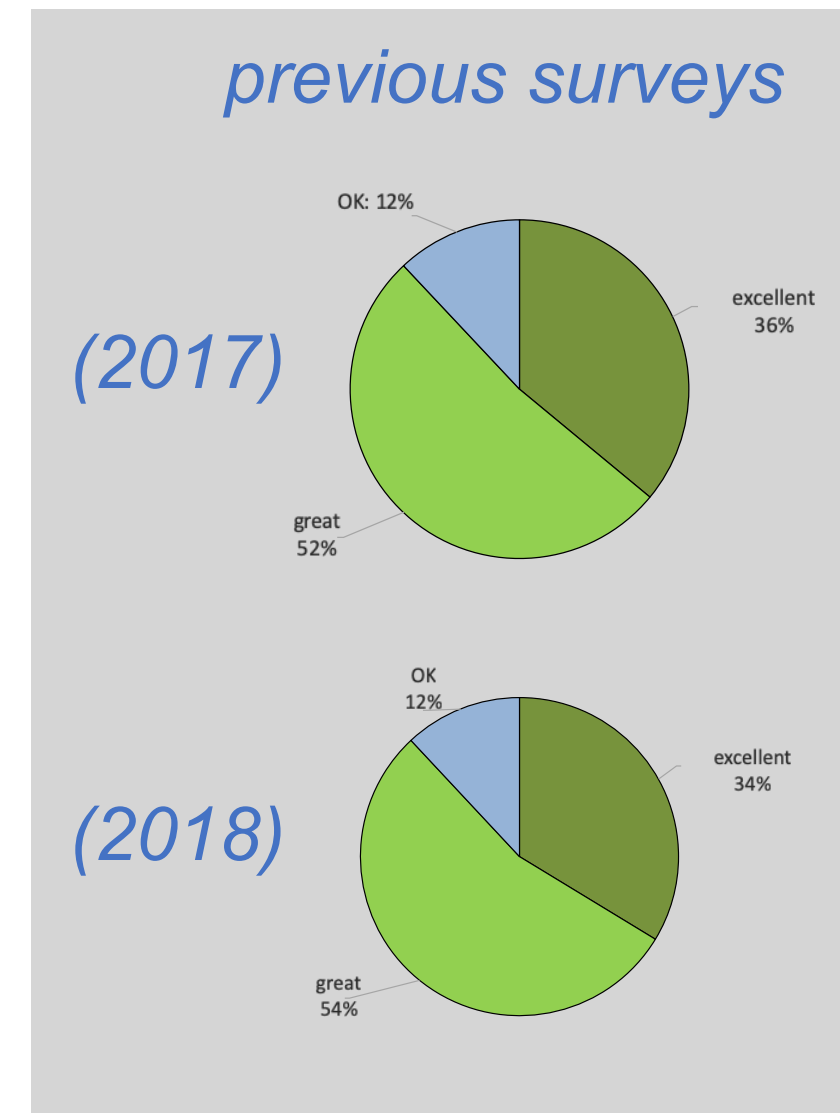
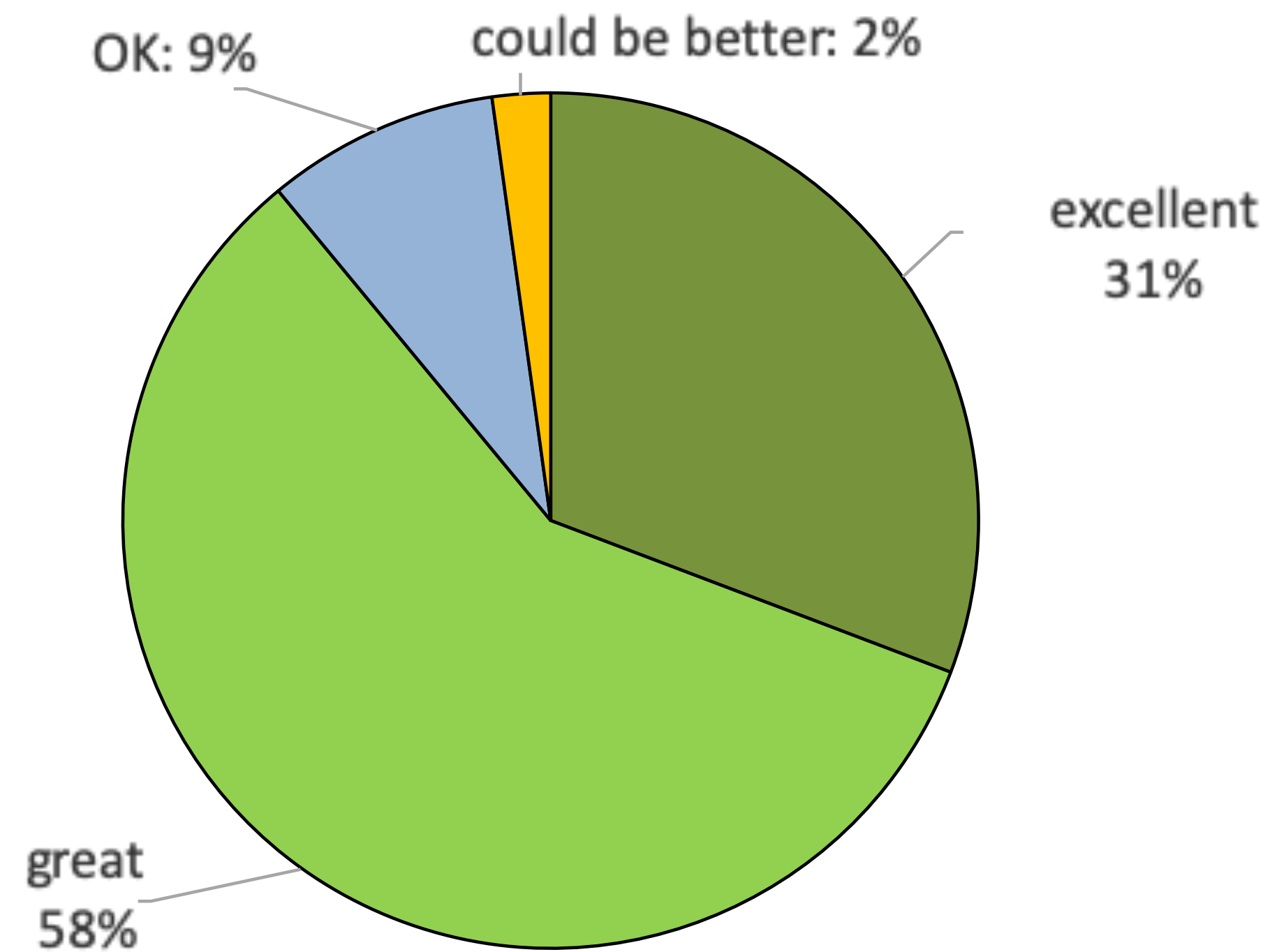
Overview of unique contributors (per year)



- ~25 unique contributors per year to EasyBuild framework
- 30-35 unique contributors per year to easyblocks, fairly stable since 2015
- 80-90 unique contributors per year for easyconfigs

3rd easybuild User Survey

How would you rate the overall quality of EasyBuild?



98% positive, great + excellent slightly increasing... (88% ➡ 91%)

(or even 100% positive if you ignore the two Spack developers who participated in the EasyBuild survey)

(nobody answered "pretty bad")

(very similar to previous survey)

Goals for 2020

- **keeping up with incoming contributions/pull requests**
 - we will break 10,000 easyconfig PRs soon...
 - soon over 2000 supported software packages (excl. extensions)
 - more automation? semi-automatic easyconfig PR tests?
 - you can help as a contributor... (see Sam's tutorial)
- **improved EasyBuild regression test (+ public dashboard?)**
 - using ReFrame
 - using container images to control test environment (OS, installed packages, ...)
 - using QEMU to test on multiple processor architectures?
- **better error reporting**

Challenges in 2020 (1/2)

- insatiable hunger of scientists for getting more software packages installed...
- maintaining large bundles of extensions (see for example R easyconfigs)
 - + dealing with changes checksums (R packages in CRAN)
- complete switch away from Travis CI to GitHub Actions?
 - @boegelbot currently only keeps an eye on Travis...
- maintain compatibility with Python 2 + Python 3.5, 3.6, 3.7, 3.8, etc.
 - we should anticipate (minor) breaking changes in every upcoming Python release...

Challenges in 2020 (2/2)

- death of Python 2 (how to deal with scientific software that still requires it)
- growing variety in system architecture: x86_64 (Intel, AMD), POWER, ARM, ...
 - rise of AMD Rome => requires GCC 9.x, optimal BLAS/LAPACK library?
 - impact on common toolchains?
- Intel oneAPI
 - will replace Parallel Studio in 2021 (?)
 - changed compiler commands (?), filesystem hierarchy, etc.

Ideas (1/2)

Support for tweaking existing easyconfig files (`eb --tweak`)

- replacement for `eb --try-*` (in the long term)
- only generate new easyconfigs (don't also go ahead and install them too)
- example use case: update toolchain + dependencies to *specified* versions

```
$ eb example-1.2-foss-2020a.eb --tweak version=1.3
```

```
$ eb example.eb --tweak toolchain=foss:2020b
```

```
$ eb example.eb --tweak deps=Python:3.9.1
```

```
$ eb example.eb --tweak deps=Python:3.9.1
```

```
$ eb TF.eb --tweak version=1.2 toolchain=foss:2020b --robot
```

```
$ eb R.eb --tweak version=3.8.0 exts=latest
```

NOT IMPLEMENTED (YET)

Ideas (2/2)

- support for auto-updating to latest available software versions
 - query upstream (GitHub, PyPI, etc.) for available versions, pick latest
 - generate new easyconfig (+ also update dependencies if `--robot` is used)
- "smart" eb option to generate (initial) easyconfig file based on provided info

```
$ eb --new-ec ReFrame 2.21 system
```

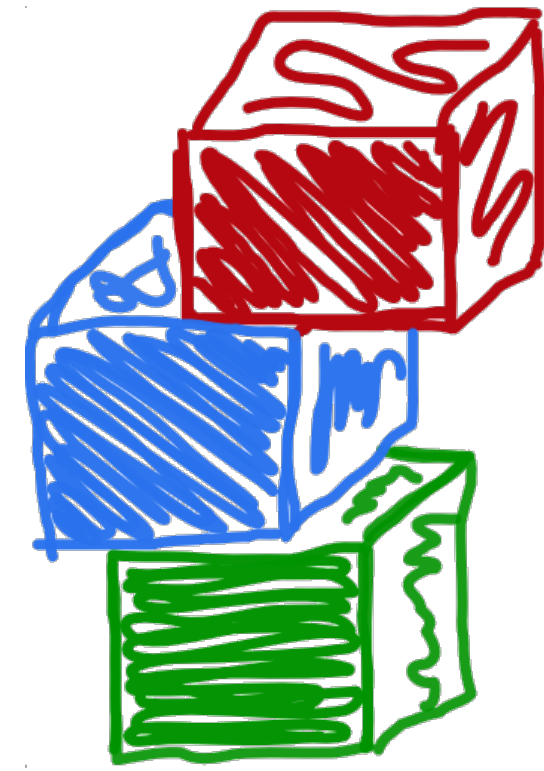
```
Created easyconfig file ReFrame-2.21.eb
```

```
$ eb --new-ec pyfoo 1.2 https://pyfoo.org deps=Python:3.9.1
```

```
    foss/2020a files=bin/pyfoo https://pyfoo.org/pyfoo-1.2.tar.gz
```

```
Created easyconfig file pyfoo-1.2-foss-2020a-Python-3.9.1.eb
```

NOT IMPLEMENTED (YET)



easybuild

Questions?

kenneth.hoste@ugent.be

easybuild@lists.ugent.be

<http://easybuilders.github.io/easybuild>

<http://easybuild.readthedocs.io>