

State of the union

Kenneth Hoste (HPC-UGent)

kenneth.hoste@ugent.be

6th EasyBuild User Meeting


virtual - Jan 25th 2021

<https://easybuild.io/eum>



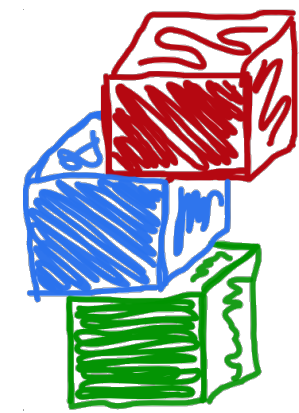
whoami

kenneth.hoste@ugent.be
@boege1 (*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  **lead developer & release manager & BDFL**
- likes family, beer, loud music, FOSS, helping people, dad jokes, stickers, ...
- doesn't like CMake, SCons, Bazel, TensorFlow, OpenFOAM, ...

- part of central IT department of Ghent University (Belgium)
- centralised scientific computing services, training & support
- for researchers of UGent, industry & knowledge institutes
- 6+1 Tier-2 clusters (~20k cores in total), ~2 PB shared storage, Infiniband, various Intel generations + AMD Rome + NVIDIA V100 GPUs (10x4)
- Tier-1 cluster "Hortense" under construction (~44k AMD Rome cores, 80 NVIDIA A100 GPUs, Infiniband)
- 8+1 team members, ~3.5k user accounts
- member of Flemish Supercomputer Centre (VSC)
<https://www.vscentrum.be>

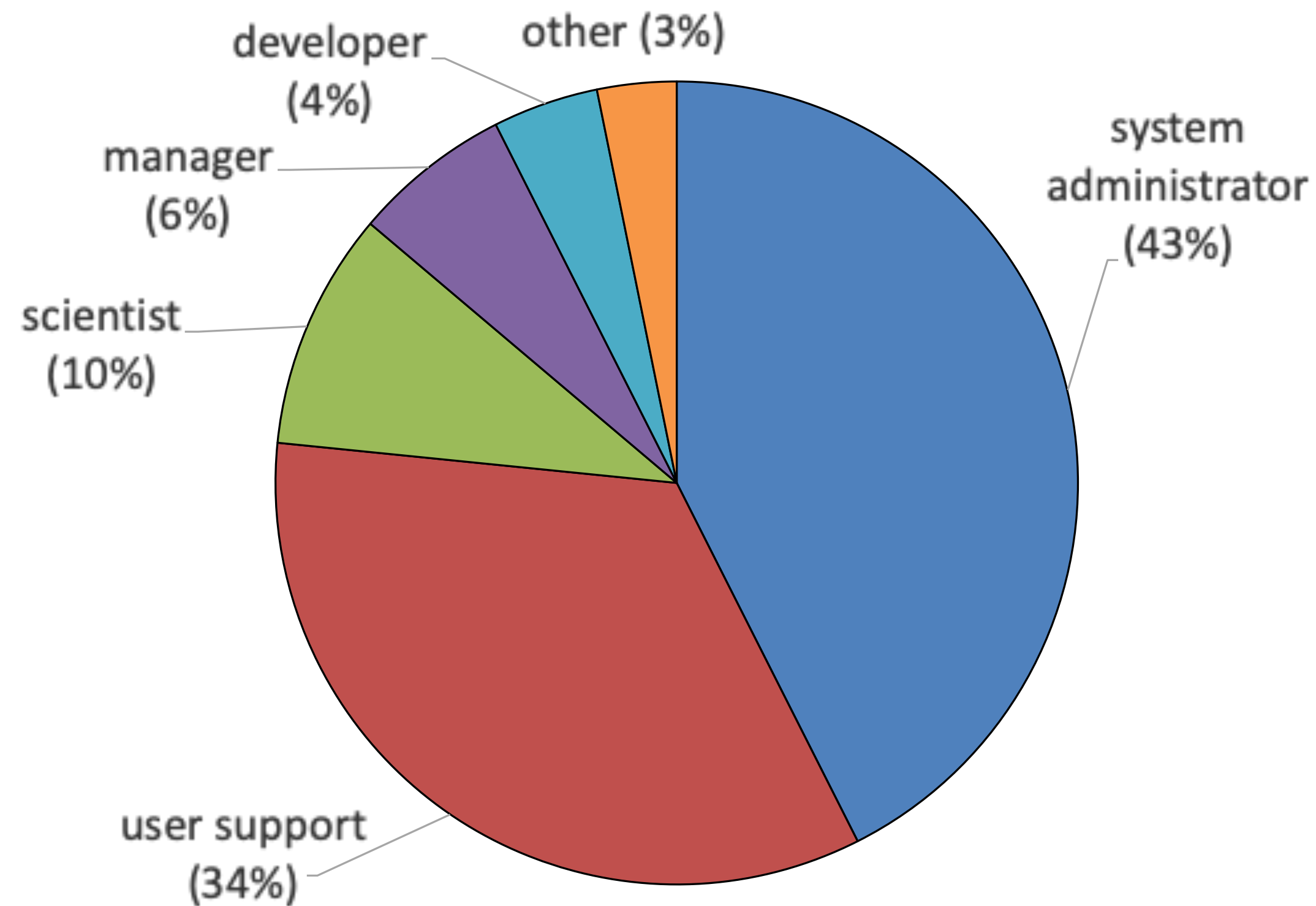




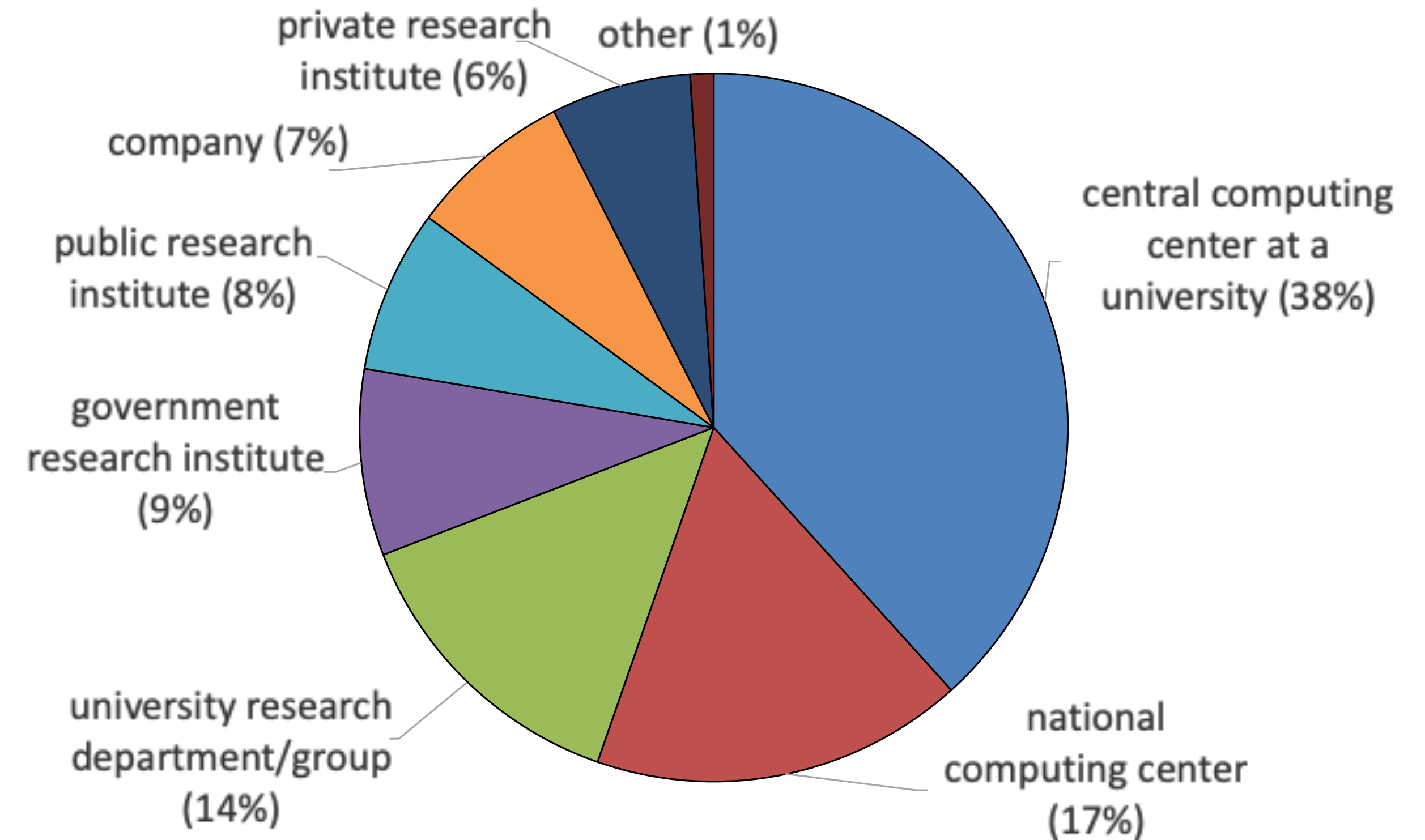
easybuild User Survey

- anonymous survey via  **SurveyMonkey**[®], 4th edition
- main objective is to get feedback from / insight into EasyBuild community
- 40 questions, covers wide variety of aspects
- **94 people participated** (77 in 2018; 93 in 2019; 91 in 2020)
- assumed to give relevant view on community
(but results are interpreted with a grain of salt)
- **full survey results (raw) are available at https://easybuild.io/user_survey**
- highlights are covered in this talk, some questions were skipped (status quo)

What is your primary profile?



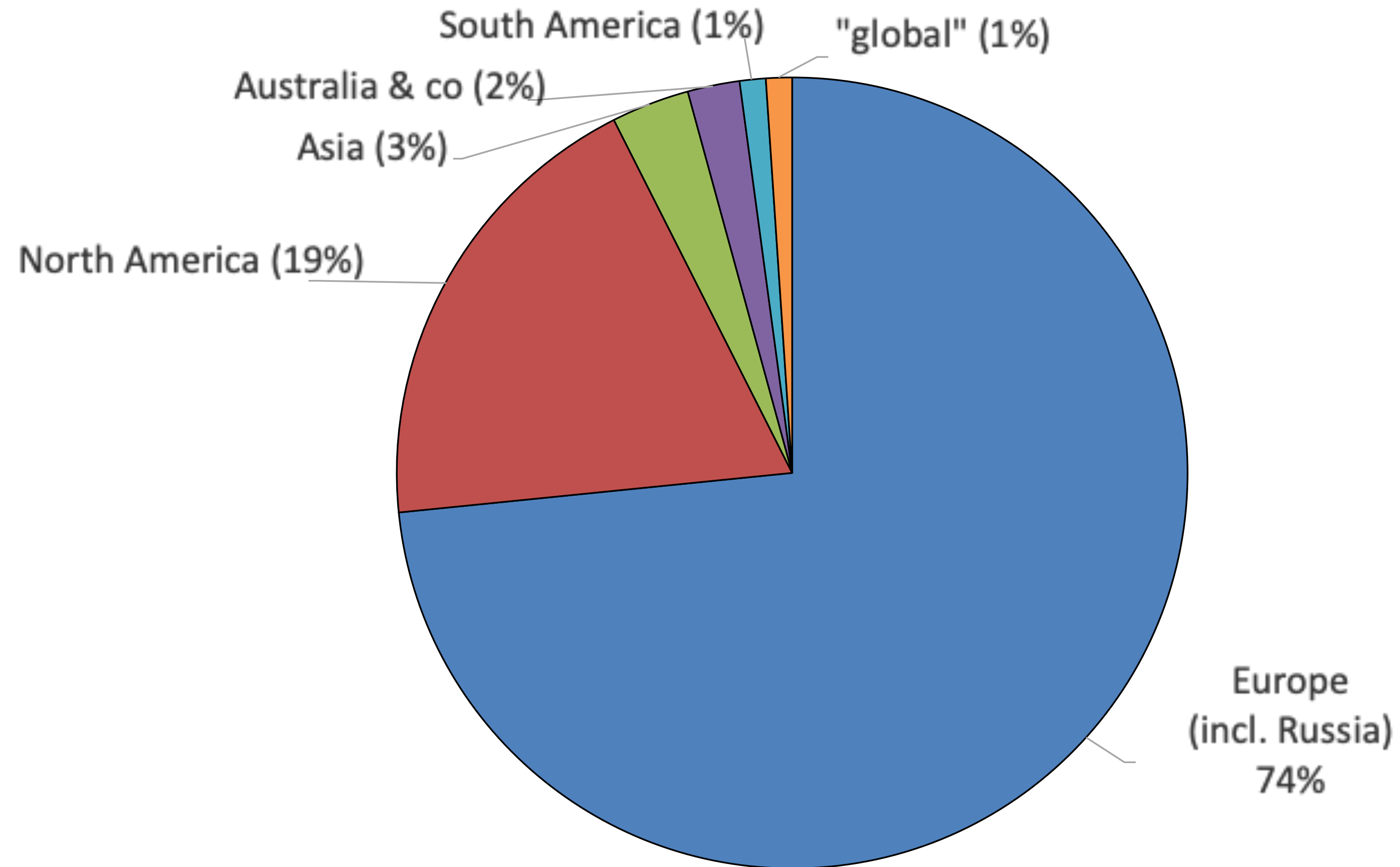
What type of organisation do you work for?



Clear bias towards sysadmins & user support.

(very similar results compared to previous surveys)

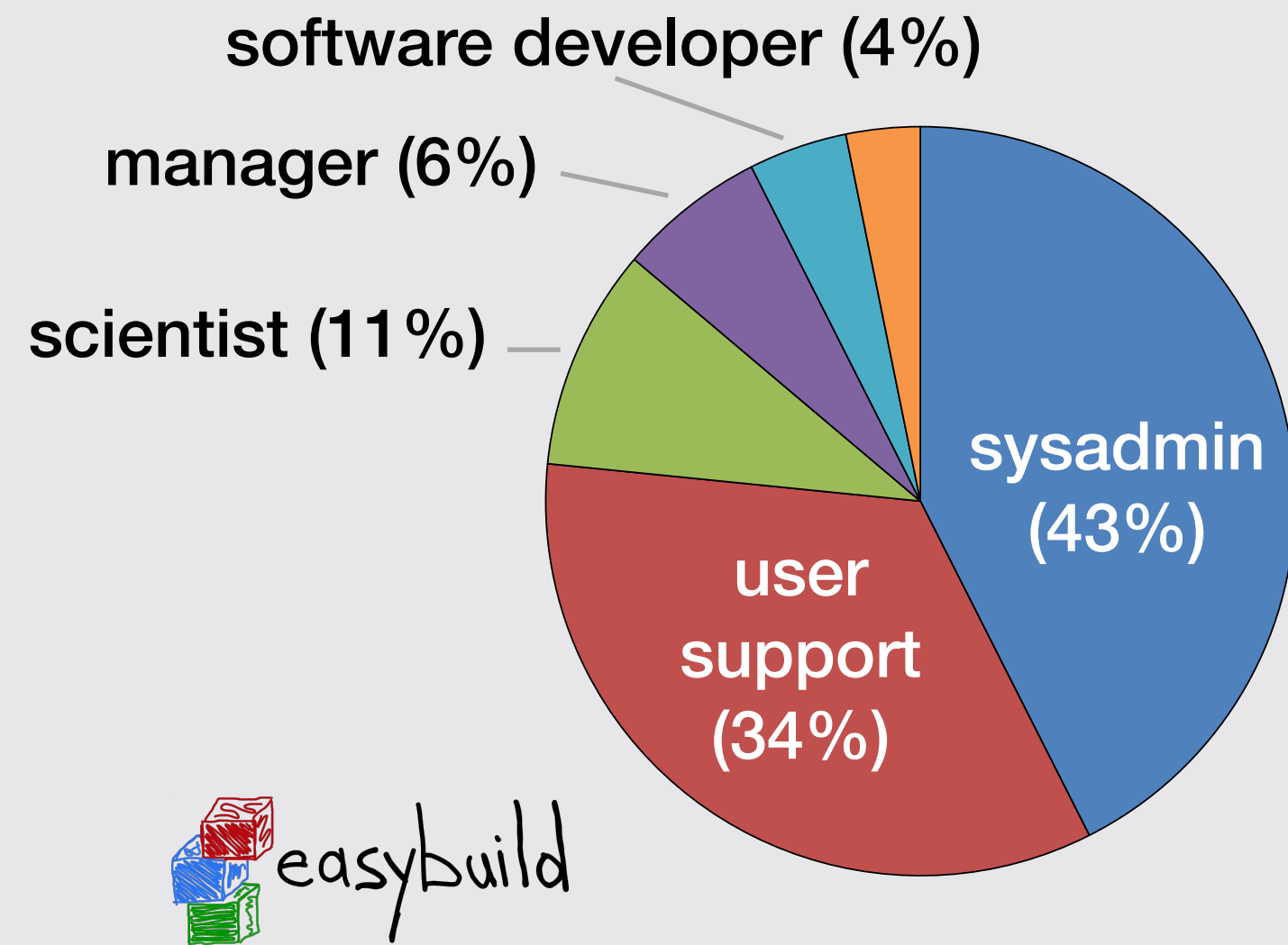
In which part of the world are you located?



(very similar results to previous surveys)

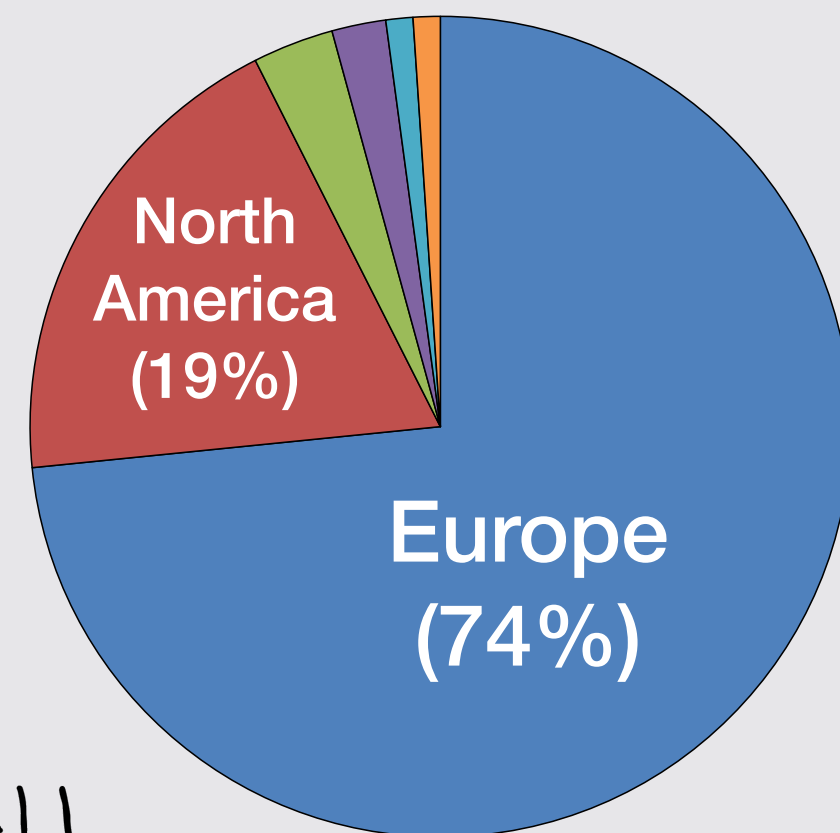
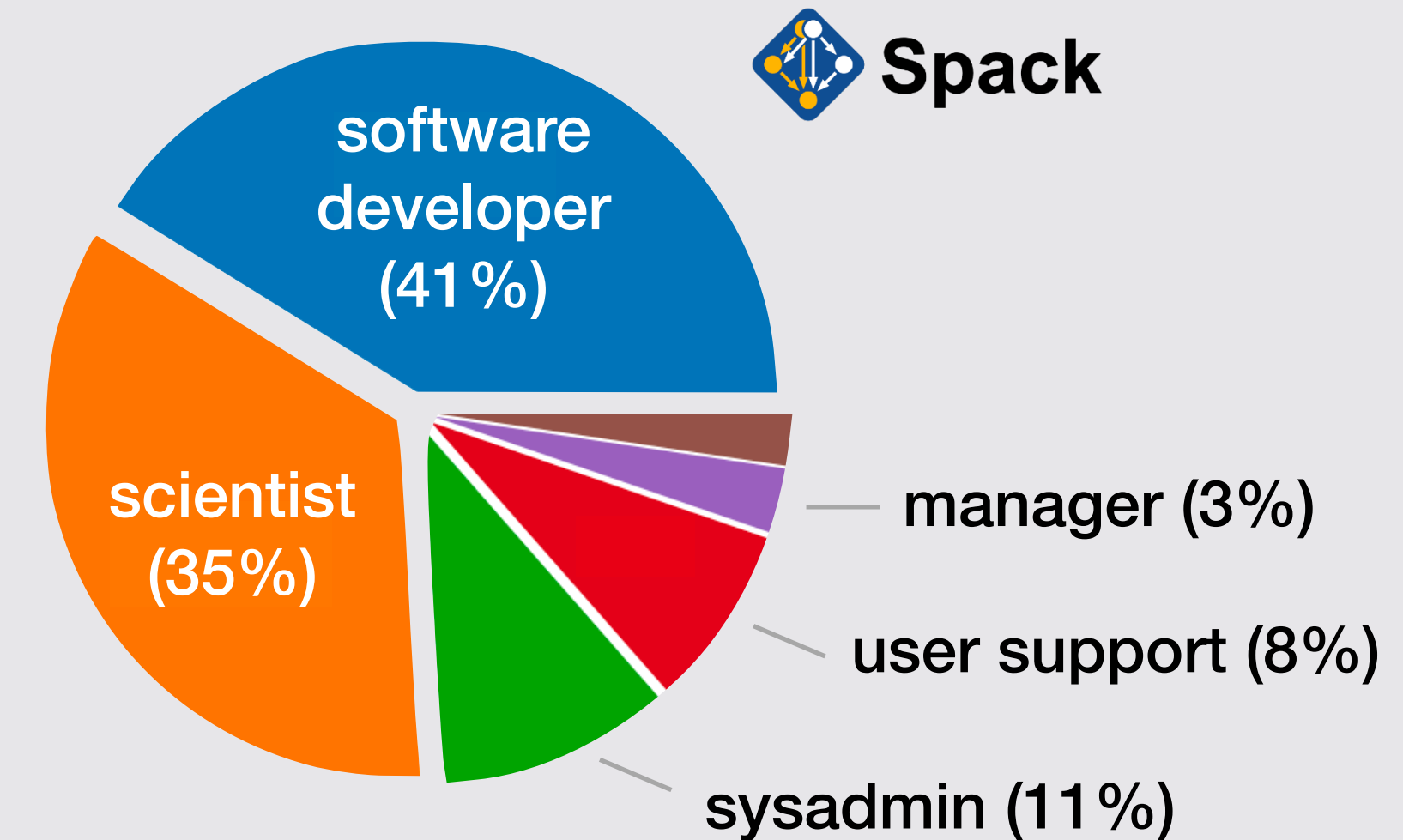
Demographics: easybuild vs Spack

(Spack data based on <https://spack.io/spack-user-survey-2020>)



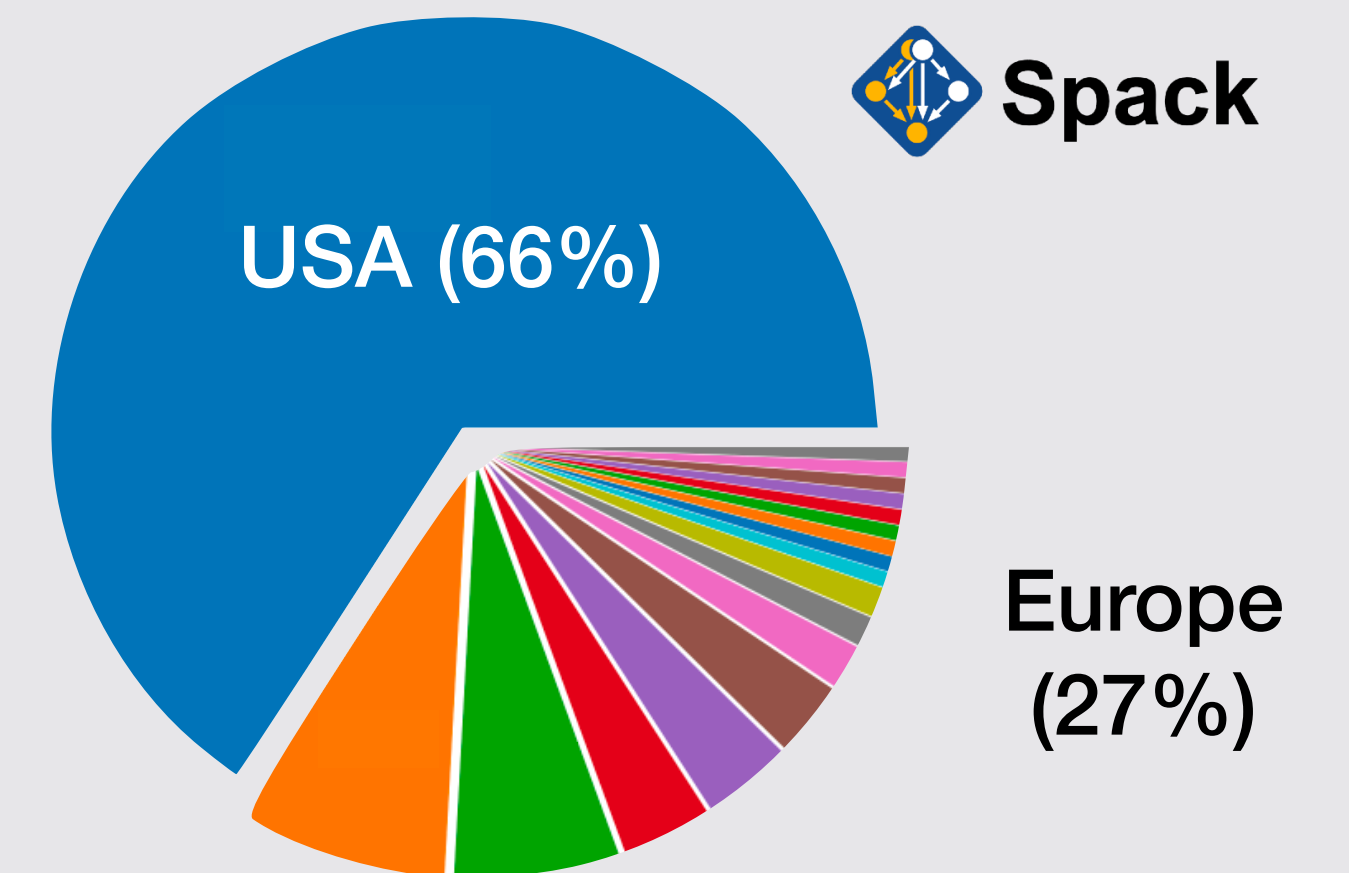
"What kind of user are you?"

- EasyBuild: sysadmins + user support (77%)
- Spack: software developers + scientists (76%)
- User communities are clearly very different!
- **Impact of different philosophies?**

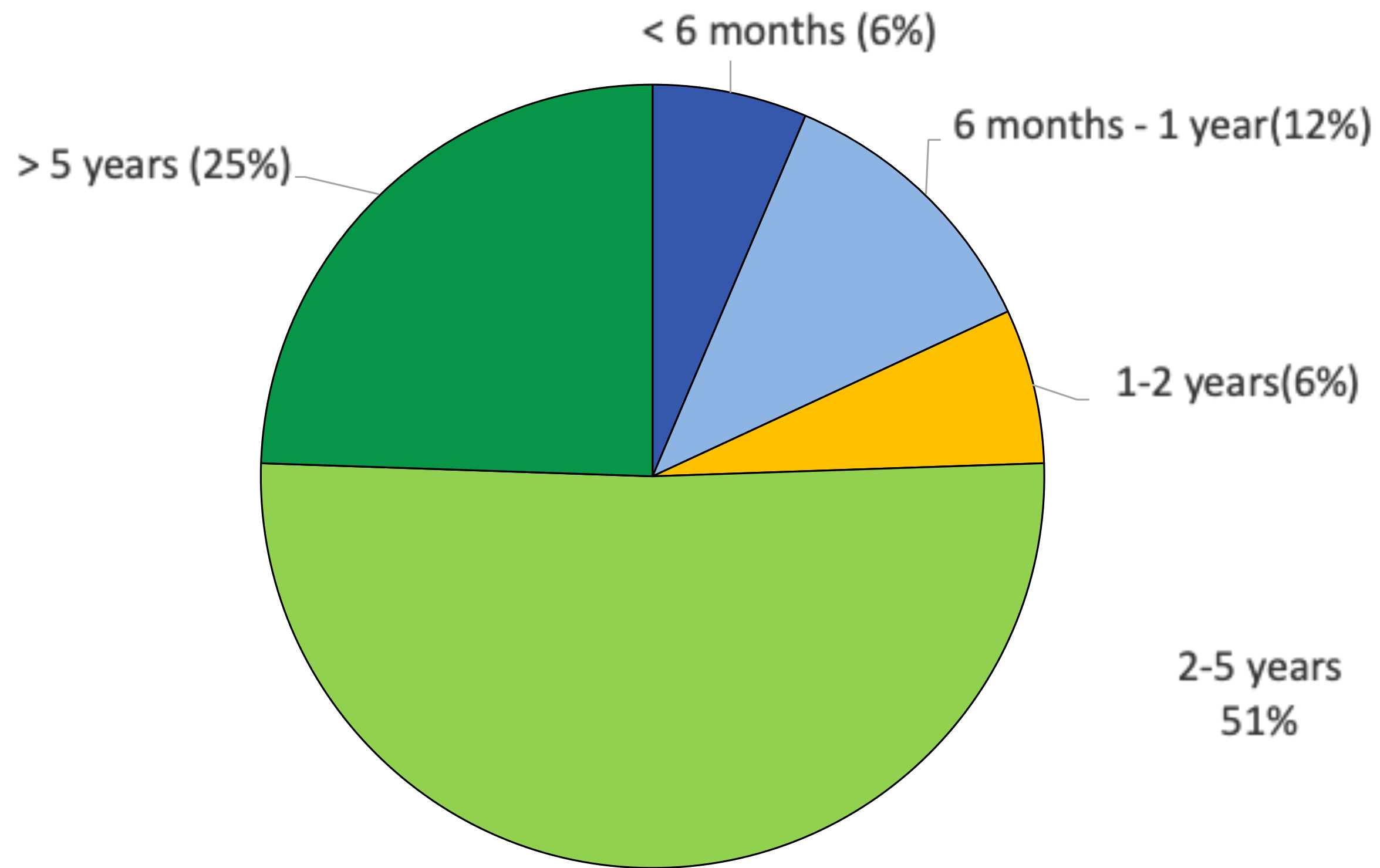


"What country are you in?"

- EasyBuild community is largely European
- Spack community is largely in US
- Can be partially explained by impact of Exascale Computing Project (ECP)



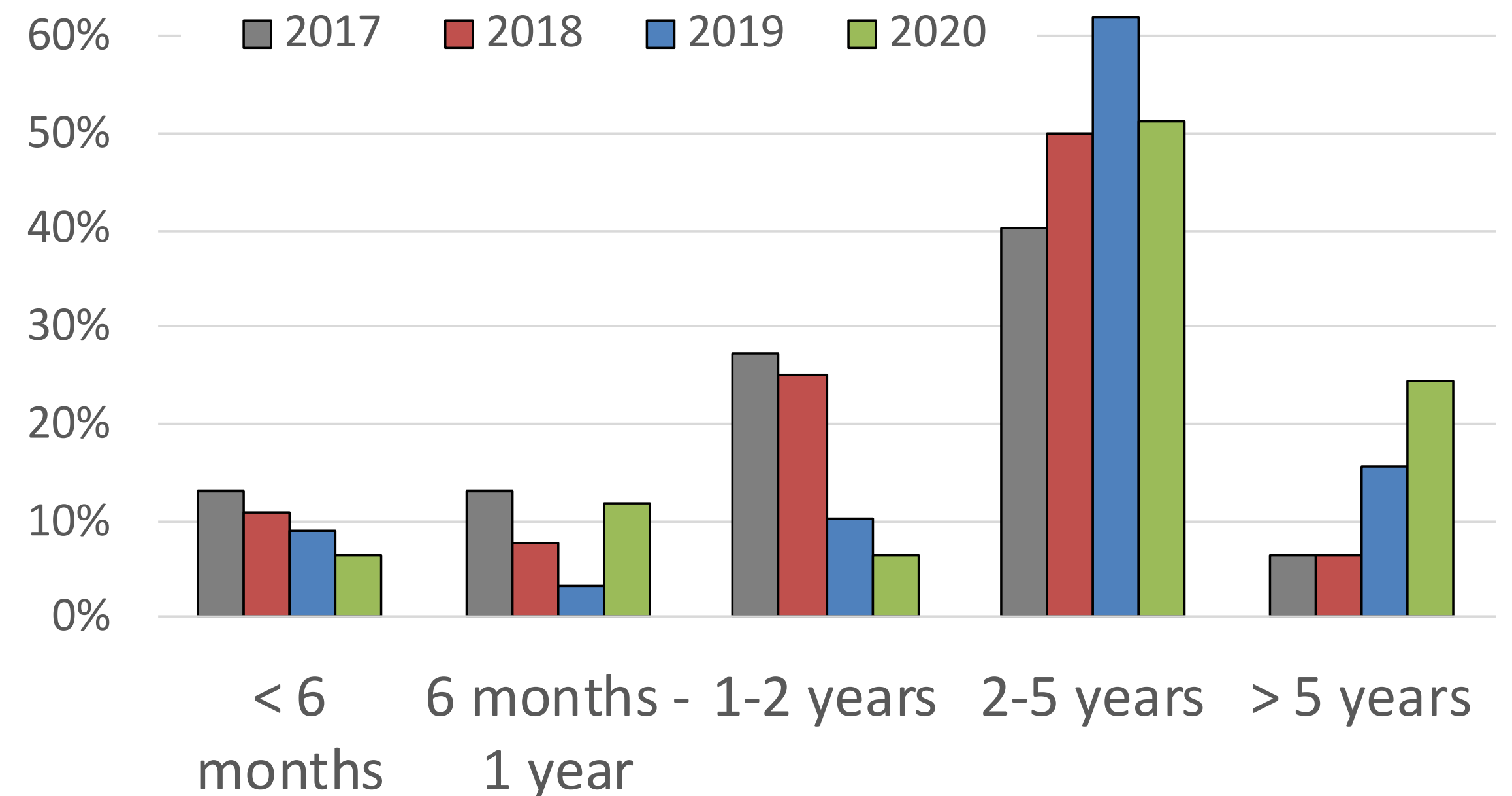
How long have you been using EasyBuild?



results of this survey (2020)

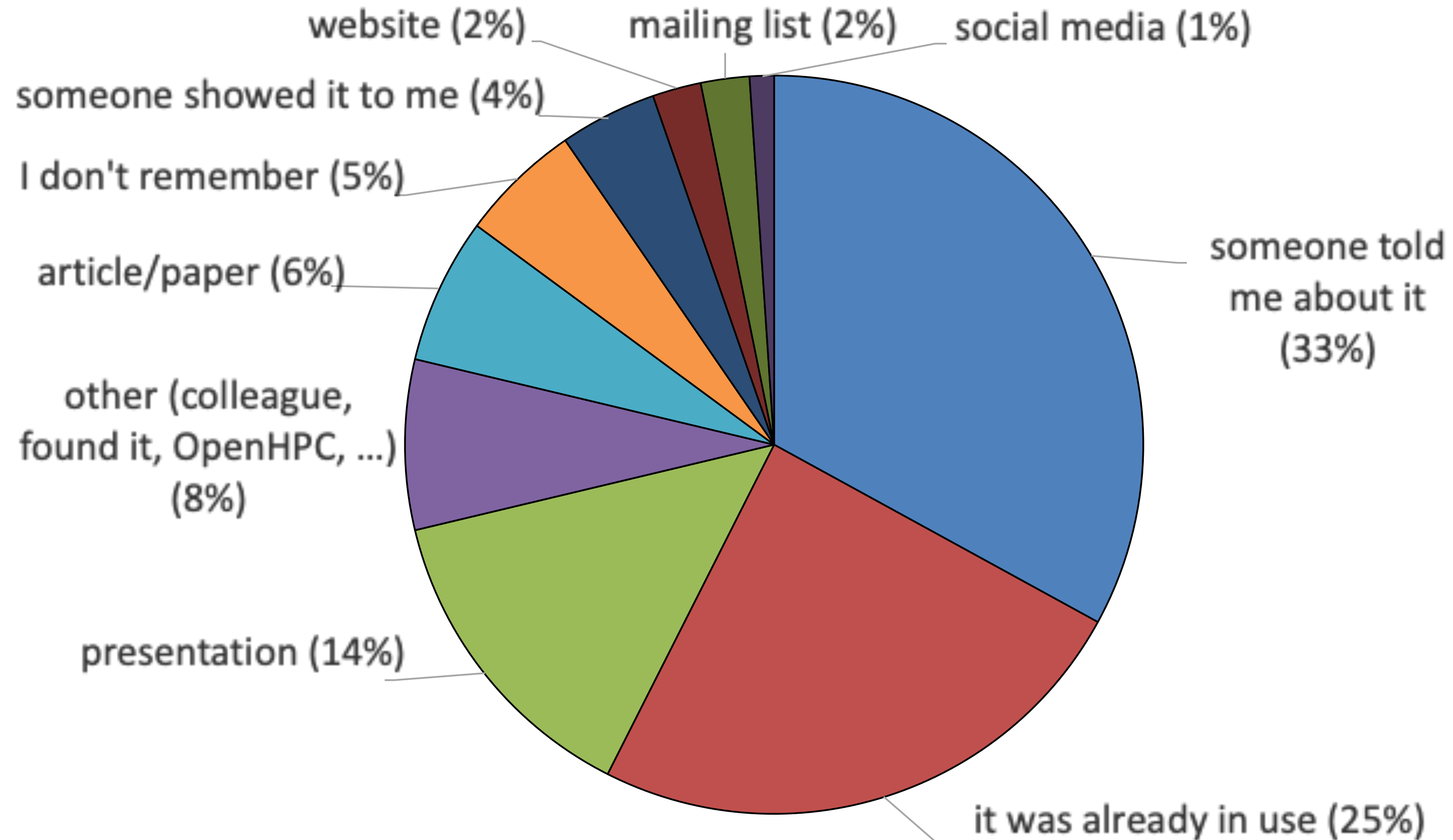
Evolution over the years:

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



easybuild User Survey

How did you first learn about EasyBuild?



Majority is still by word-of-mouth (fairly stable at ~30%),
"it was already in use" keeps increasing (4% ➡ 19% ➡ 21% ➡ 25%)

Highlights of 2020 (1/2)

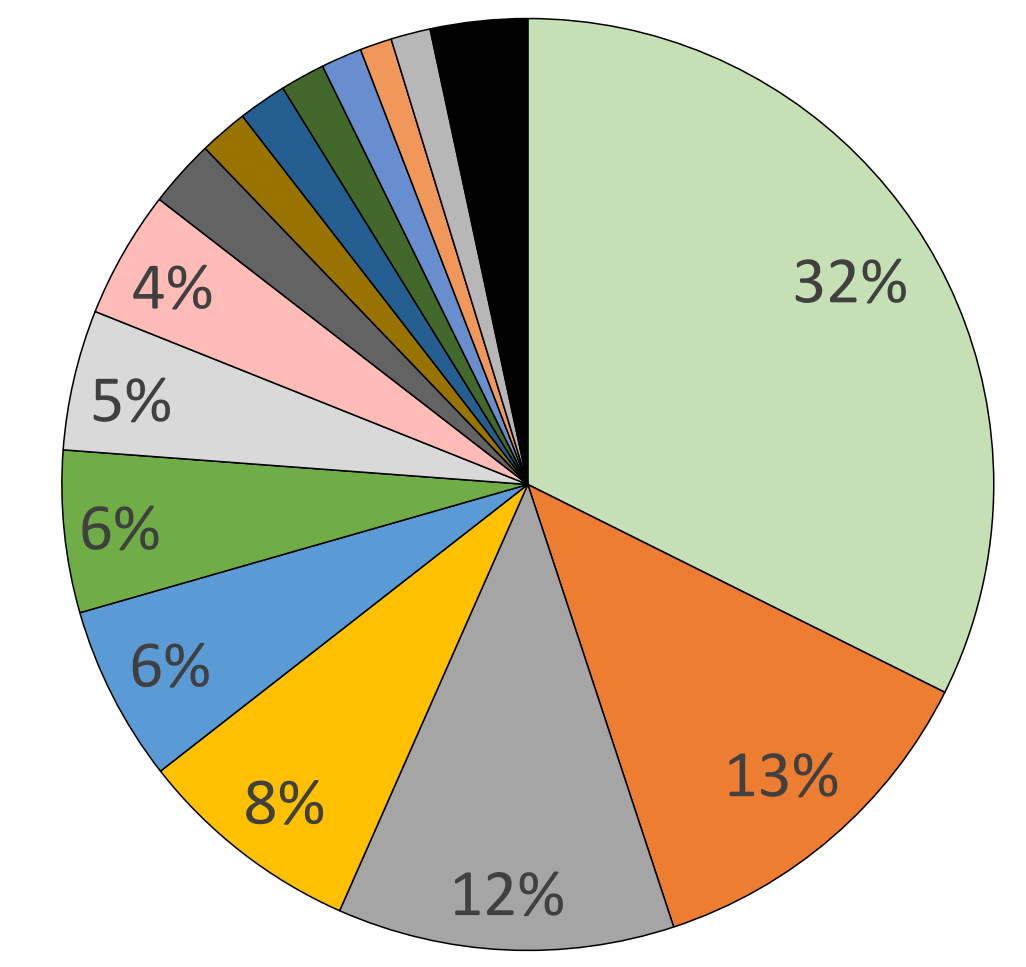
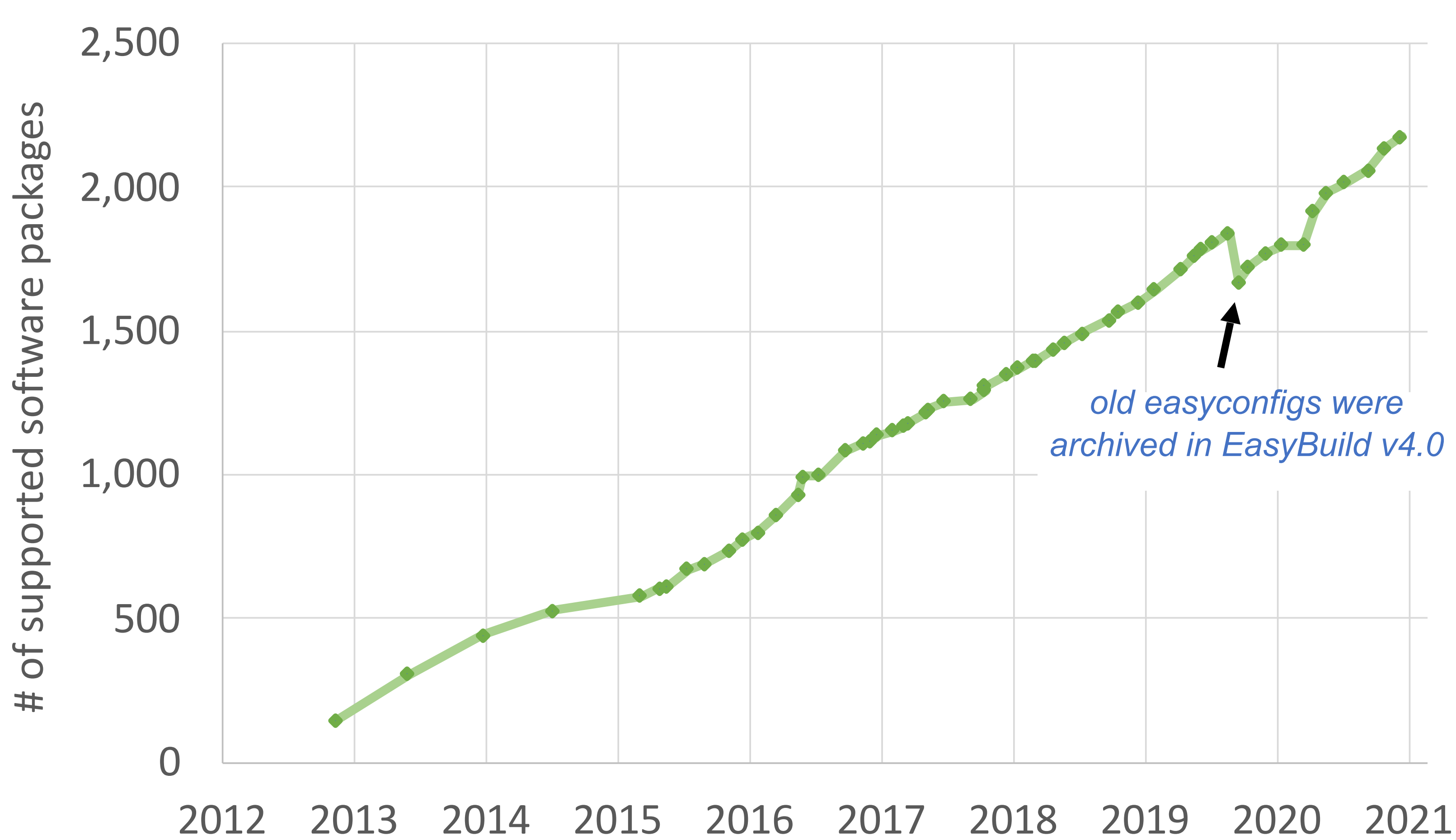


- Continued growth of EasyBuild community
- Continued growth in contributions (and managing to keep up with them)
- 23% increase in number of supported software packages
- **Proper EasyBuild tutorial** \o/
- Our very first CVE! <https://nvd.nist.gov/vuln/detail/CVE-2020-5262>
 - GitHub token was leaking in EasyBuild log in specific setting, fixed in EasyBuild v4.1.2
- Phasing out the use of Travis CI for running test suites (now GitHub Actions)
- Making all Python code PEP8 complaint + include code style checking in CI tests
- New domain: <https://easybuild.io>

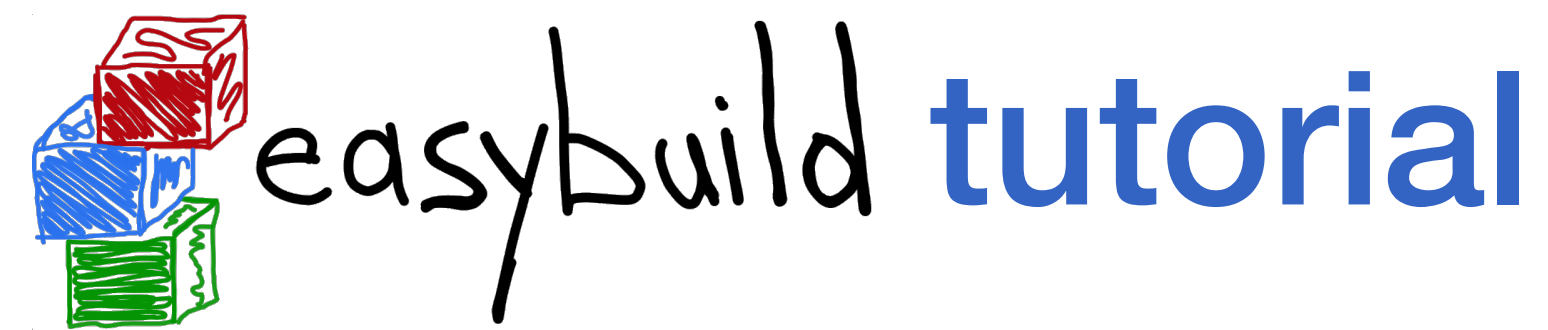
Supported software



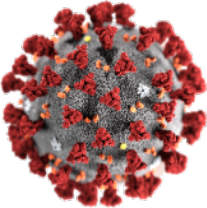
Evolution of number of supported software packages (without extensions)



- 1/3rd of supported software is bioinformatics
- 23% is (generic) libraries & tools
- Towards 2,500 different supported software packages (excl. versions)...



<https://easybuilders.github.io/easybuild-tutorial>

- Proposal for EasyBuild tutorial was accepted for ISC'20 !
- ... but all tutorials were postponed until ISC'21 due to COVID-19 pandemic 
- So we organised an online EasyBuild tutorial ourselves (via YouTube + Slack)
- **Over 100 registrations!**
- A bit too ambitious w.r.t. amount of content and time that was planned to cover it...
- Nevertheless, very positive evaluations: 4.3 (out of 5) average score!
- **We will tweak this tutorial for ISC'21, stay tuned for more info!**

A proper home for

- New domain: **<https://easybuild.io>**
- Easy to remember EasyBuild website
(was <https://easybuilders.github.io/easybuild>)
- EasyBuild documentation available via **<https://docs.easybuild.io>**,
currently just an alias for <https://easybuild.readthedocs.io>
- Should have done this years ago...

Highlights of 2020 (2/2)



- Prominent enhancements & fixes:
 - `--try-update-deps` (experimental)
 - Locks to prevent multiple runs of the same installation at the same time
 - Various significant speedups (avoiding needless module commands or easyconfig parsing)
 - `--new-pr` for framework/easyblocks, support for `--include-easyblocks-from-pr`
 - Support for installing software in alternate system root environment (via `--sysroot`)
 - Improved RPATH wrapper script
 - **Easystack files (experimental)**
 - Significant improvements for "big" software applications like TensorFlow, PyTorch, ...
- Improved support for Arm 64-bit and POWER, as well on RHEL8-based OSs



Easystack files

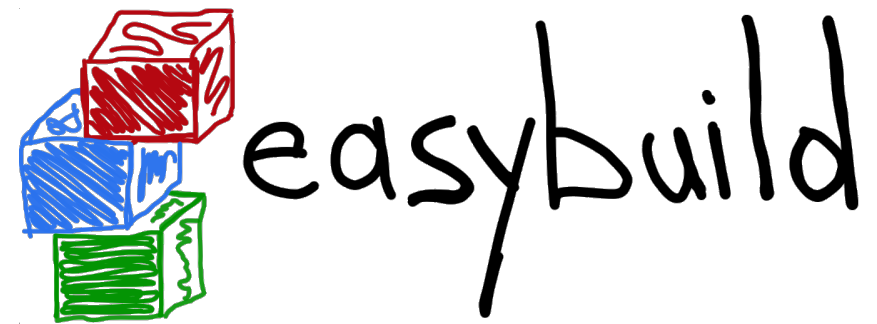


<https://docs.easybuild.io/en/latest/Easystack-files.html>

Example:

- Experimental support for easystack files was added in EasyBuild v4.3.2
- Describes software stack to install (YAML syntax)
- Primary key: software, then toolchains, software versions
- Enhancements to be implemented:
 - Support for including EasyBuild configuration settings
 - Filtering based on "labels" in EasyBuild configuration
 - Support for using regular expressions

```
software:
  Bioconductor:
    toolchains:
      foss-2020a:
        versions:
          3.11:
EasyBuild:
  toolchains:
    SYSTEM:
      versions: [4.3.1]
GROMACS:
  toolchains:
    foss-2020a:
      versions:
        2020.1:
        2020.3:
    fosscuda-2020a:
      versions: [2020.1]
OpenFOAM:
  toolchains:
    foss-2020a:
      versions: [8, v2006]
R:
  toolchains:
    foss-2020a:
      versions: [4.0.0]
```



From EasyBuild to EESSI

<https://www.eessi-hpc.org>

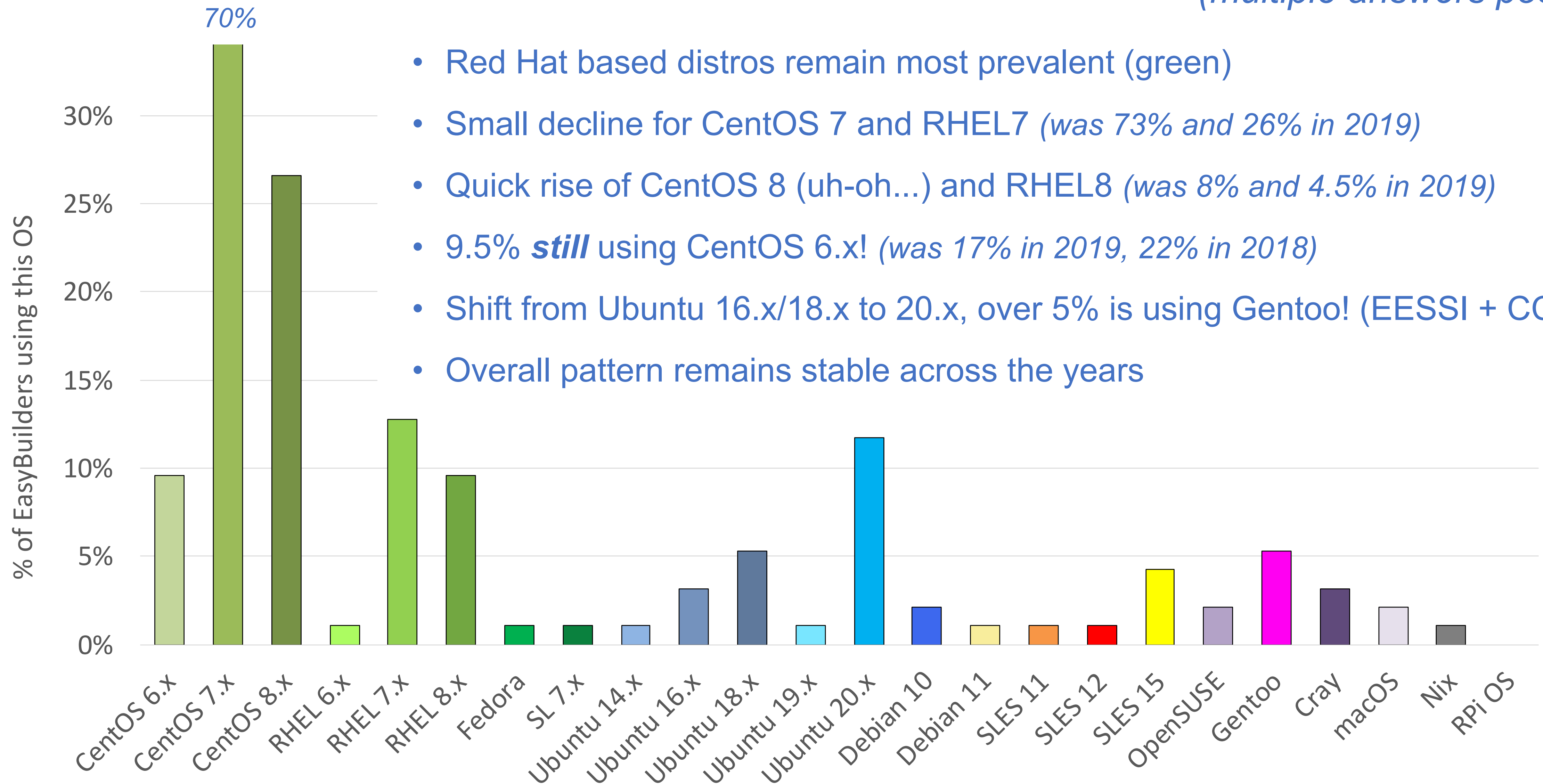
<https://eessi.github.io/docs>



- European Environment for Scientific Software Installations (EESSI)
- **Goal: building a *shared* repository of scientific software installations**
- Combines various open source projects, including CernVM-FS, Gentoo Prefix, EasyBuild, Lmod, archspec, etc.
- In some sense, EESSI is "next-gen EasyBuild": working together on the installations themselves, rather than (only) on a tool for automating them...
- **See Bob's EUM talk on the EESSI project on Tuesday at 12:00 UTC!**

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

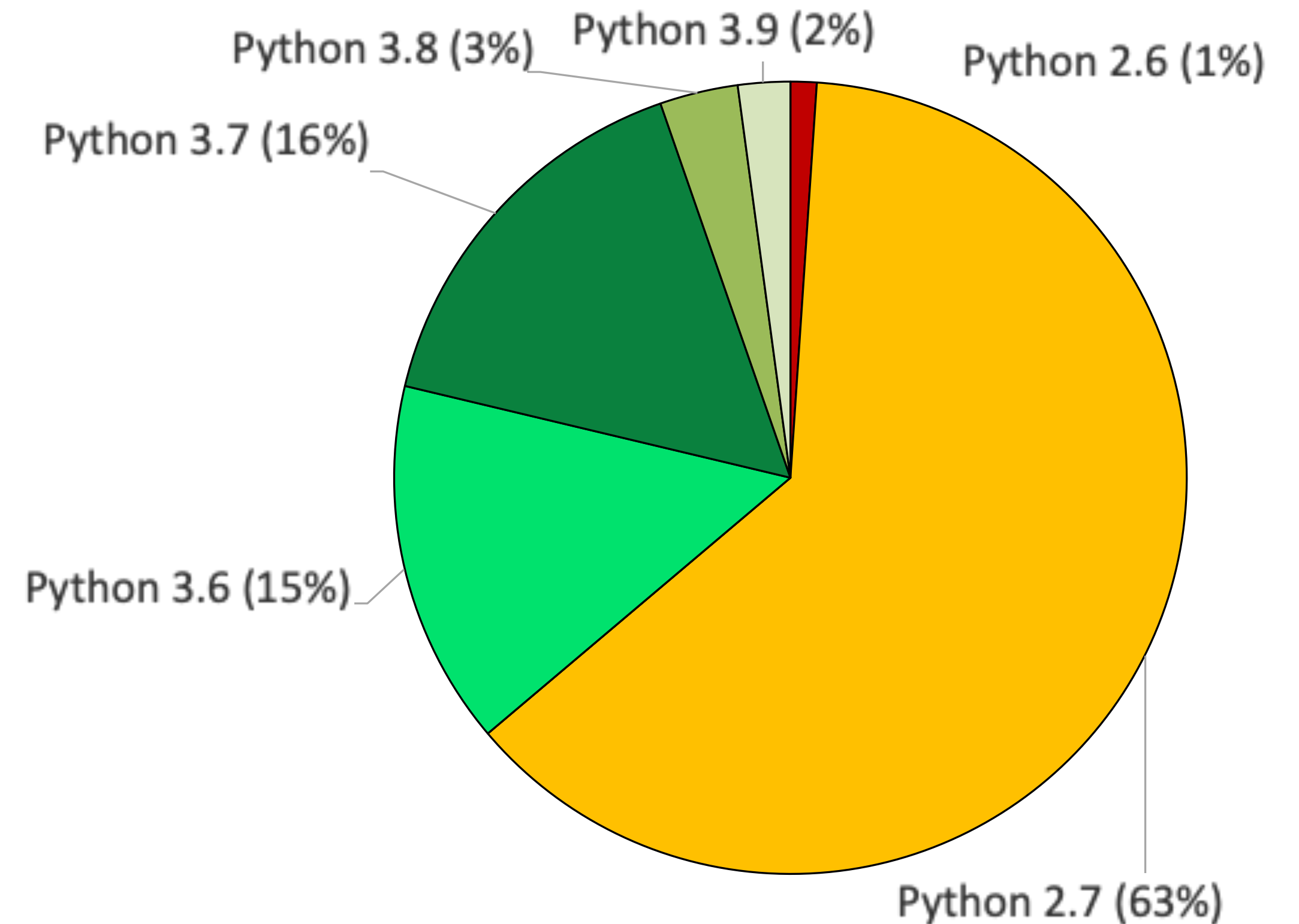
(multiple answers possible)



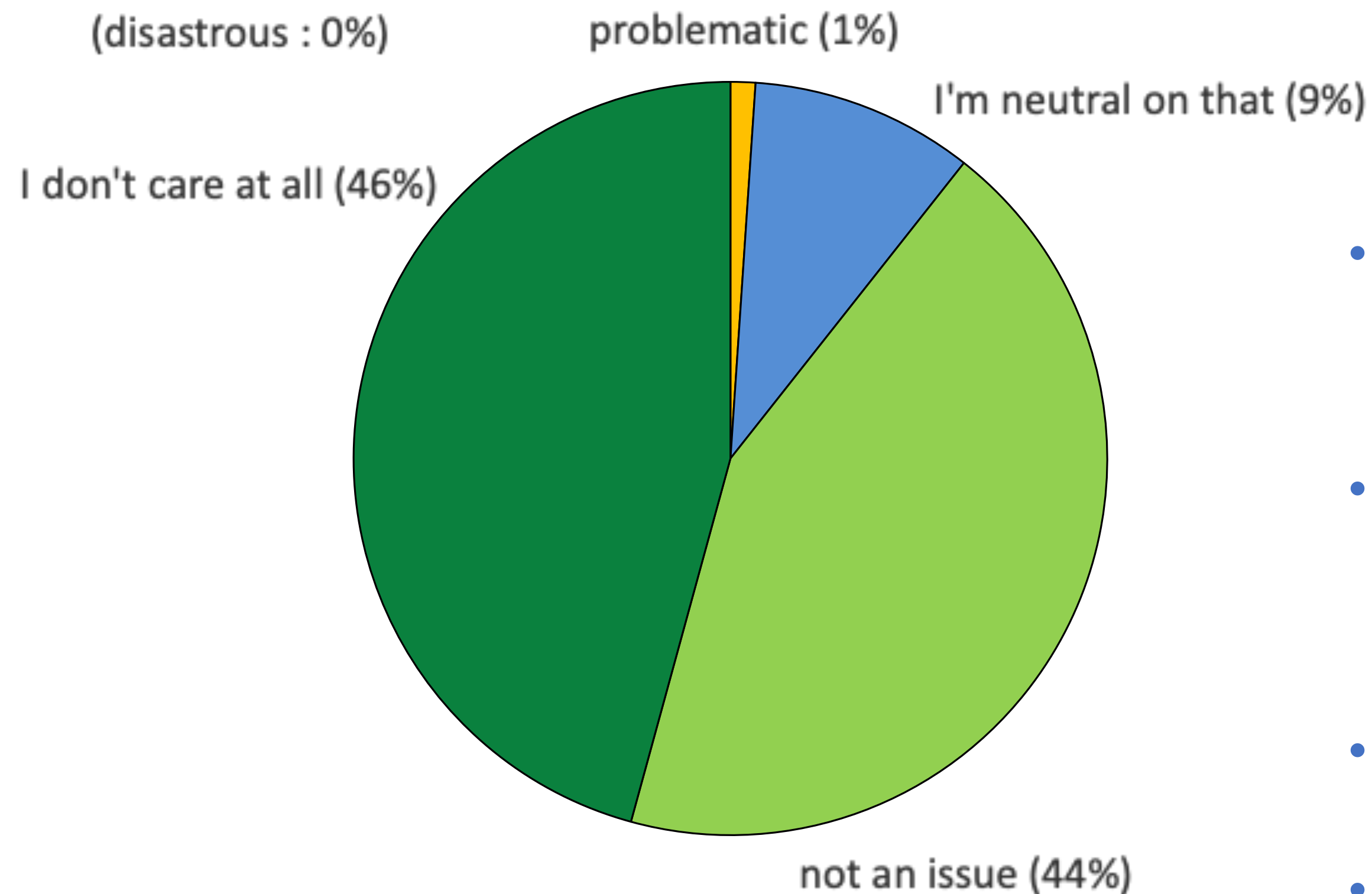
- Red Hat based distros remain most prevalent (green)
- Small decline for CentOS 7 and RHEL7 *(was 73% and 26% in 2019)*
- Quick rise of CentOS 8 (uh-oh...) and RHEL8 *(was 8% and 4.5% in 2019)*
- 9.5% **still** using CentOS 6.x! *(was 17% in 2019, 22% in 2018)*
- Shift from Ubuntu 16.x/18.x to 20.x, over 5% is using Gentoo! (EESSI + CC?)
- Overall pattern remains stable across the years

Which Python version do you usually use to run EasyBuild?

- Python 2.6 just about dead... (4% → 1%)
- Markable decline for Python 2.7 (80% → 63%)
- Adoption of Python 3.x is still slow-ish... (16% → 36%)
- splintered use across different Python 3 versions...
- my predictions from previous EUM:
 - ✓ majority still using Python 2 (CentOS 7/RHEL 7)
 - ✓ significant increase in use of Python 3
 - ✗ Python 3.6 is most prevalent Python 3 version (so close!)
- Predictions for 2021:
 - (small) majority still using Python 2 due to CentOS/RHEL 7.x (but it will be close!)
 - More splintering across different Python 3.x versions
 - Python 3.8 is most prevalent Python 3 version

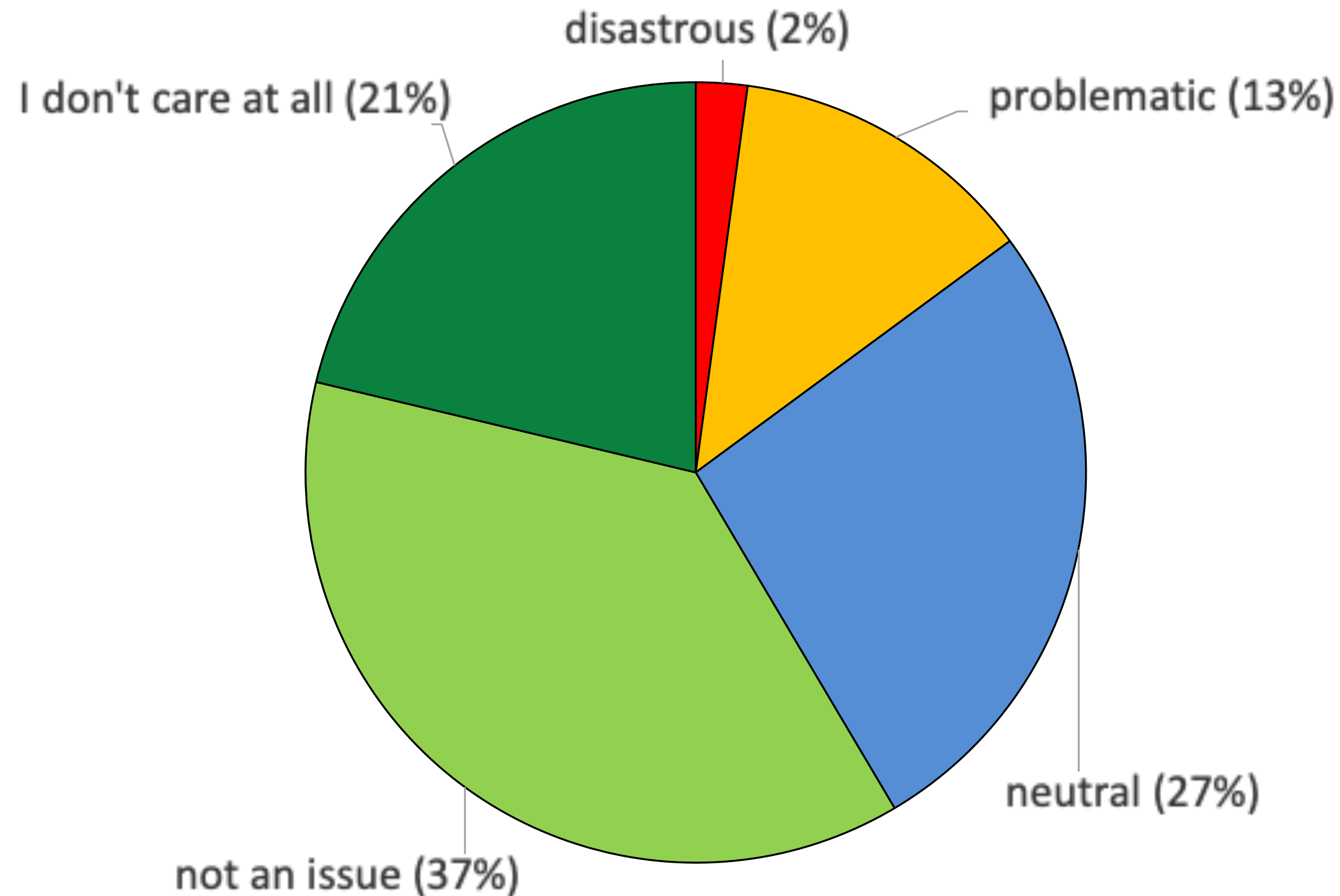


How troublesome would it be for you if EasyBuild becomes incompatible with Python 2.6?

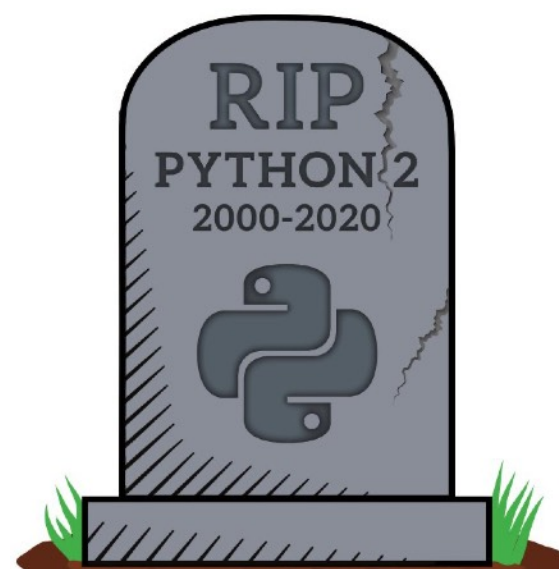


- Support for running EasyBuild on top of Python 2.6 was deprecated in EasyBuild 4.0 (Sept'19)
- Declining number of people care about Python 2.6 (13% 🗑️ 4% 🗑️ 2% 🗑️ 1%)
- People stuck on CentOS/RHEL 6 have adopted Python 2.7
- It seems like the world is finally ready to let Python 2.6 die...

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



- Still some concern about dropping Python 2 support (25%  15%)
- **no plans currently to drop Python 2 support**
- keeping Python 2.7 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 (?)



Using Python 3 to run EasyBuild



<https://docs.easybuild.io/en/latest/Python-2-3-compatibility.html>

- eb script determines which `python` command will be used to run EasyBuild
- By default considers `python`, `python3`, `python2` (in that order)
- You can specify which command should be used via `$EB_PYTHON`:

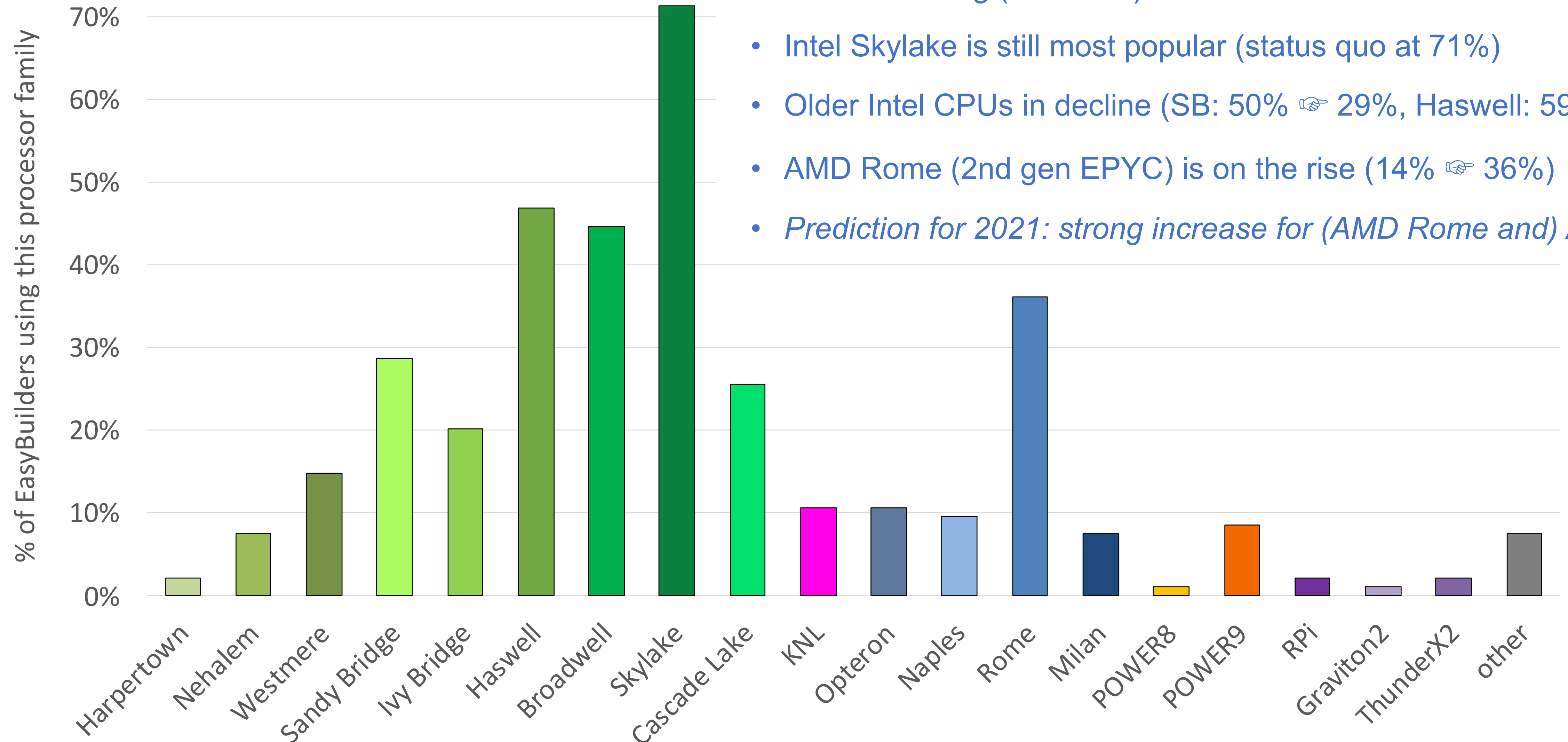
```
export EB_PYTHON=python3.6
```

```
eb --show-system-info
```

- Use this to opt-in to using Python 3 to run EasyBuild if `python` is Python 2 !
- Use "`EB_VERBOSE=1 eb ...`" to get report on what was considered...

For which processor families are you installing software using EasyBuild?

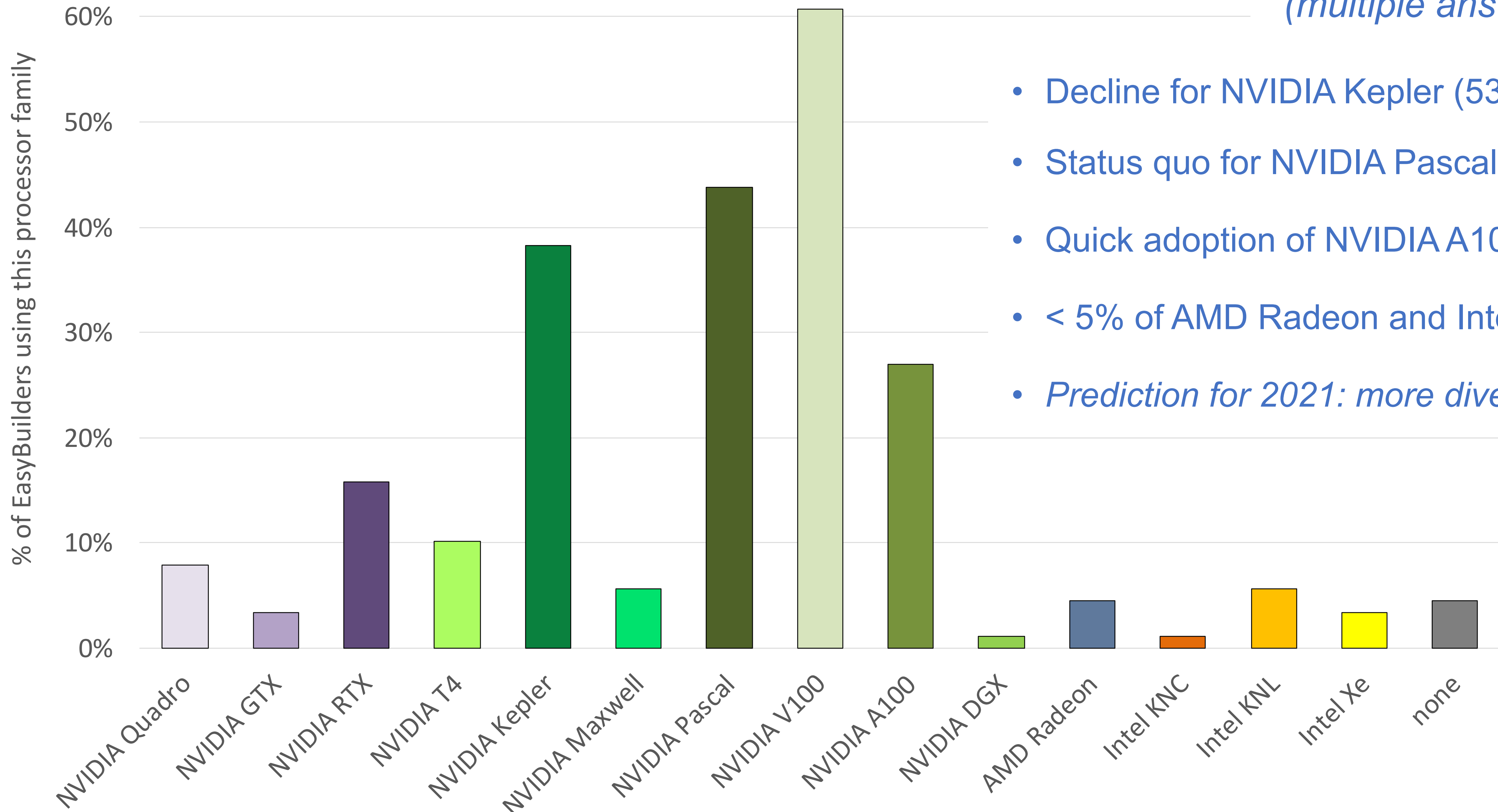
(multiple answers possible)



- Intel is still king (for now...)
- Intel Skylake is still most popular (status quo at 71%)
- Older Intel CPUs in decline (SB: 50% 📉 29%, Haswell: 59% 📉 47%)
- AMD Rome (2nd gen EPYC) is on the rise (14% 📈 36%)
- *Prediction for 2021: strong increase for (AMD Rome and) Arm!*

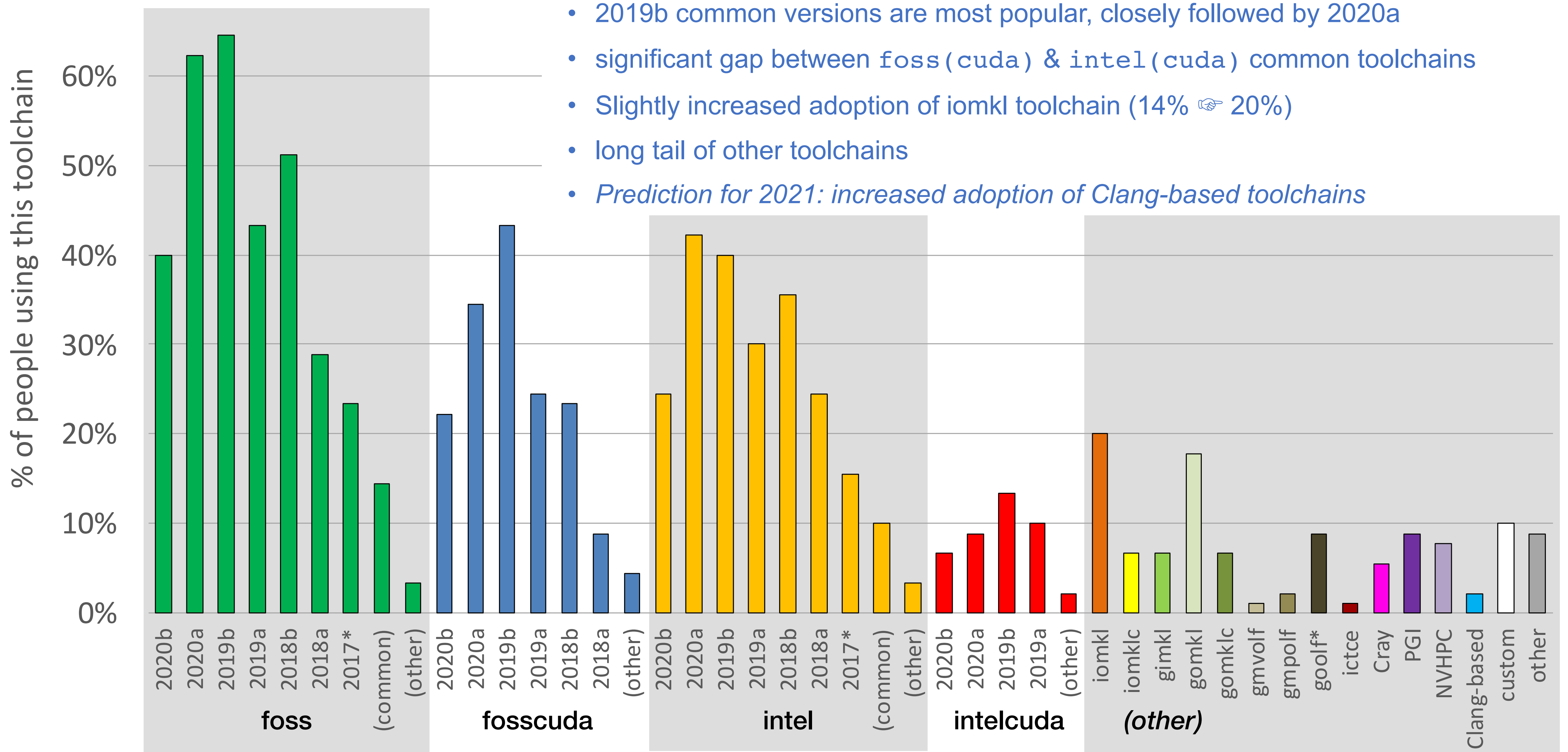
For which accelerators are you installing software using EasyBuild?

(multiple answers possible)



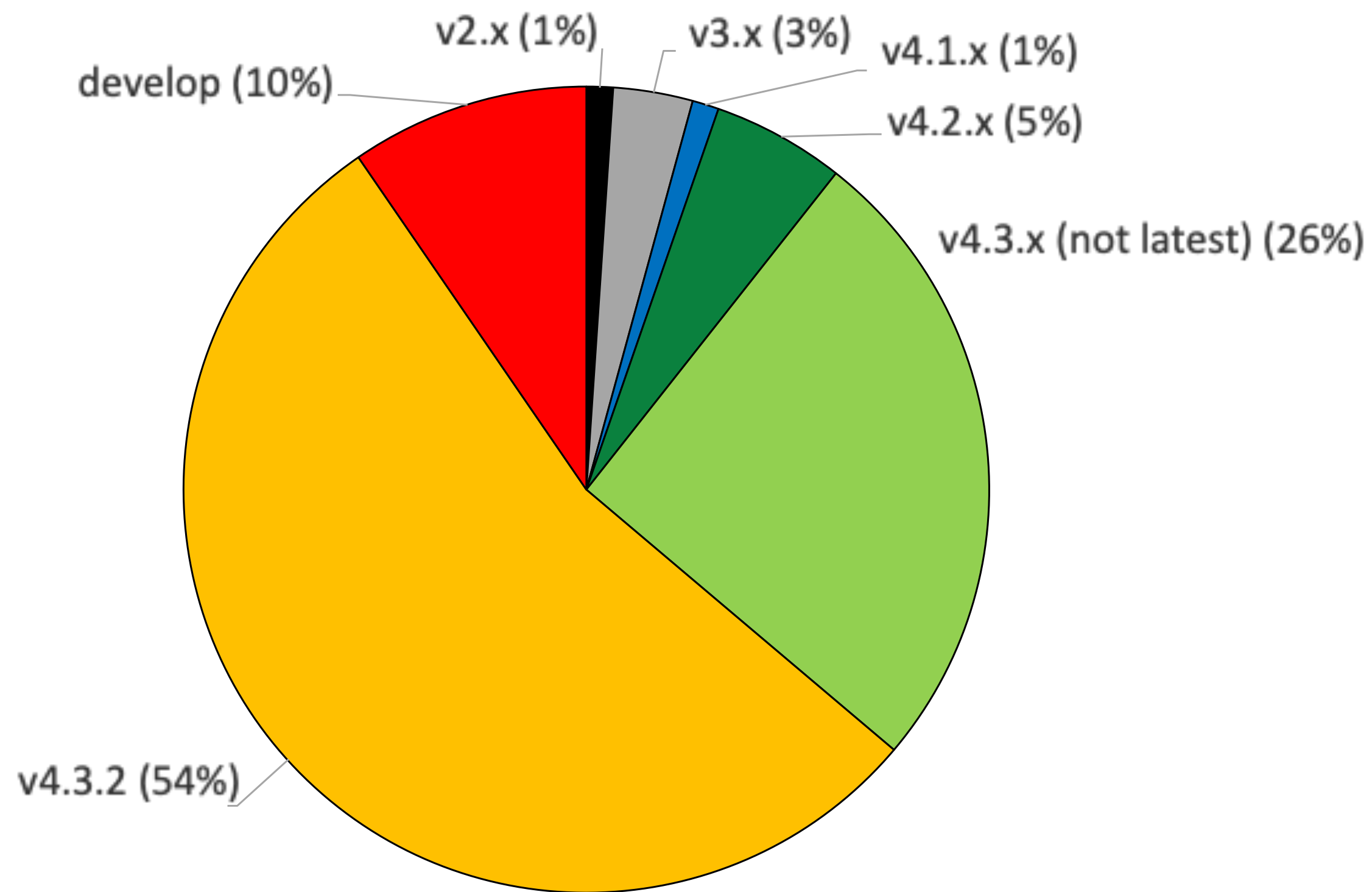
- Decline for NVIDIA Kepler (53% → 38%)
- Status quo for NVIDIA Pascal and V100
- Quick adoption of NVIDIA A100
- < 5% of AMD Radeon and Intel Xe
- *Prediction for 2021: more diversity...*

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



- 2019b common versions are most popular, closely followed by 2020a
- significant gap between foss(cuda) & intel(cuda) common toolchains
- Slightly increased adoption of iomkl toolchain (14% → 20%)
- long tail of other toolchains
- *Prediction for 2021: increased adoption of Clang-based toolchains*

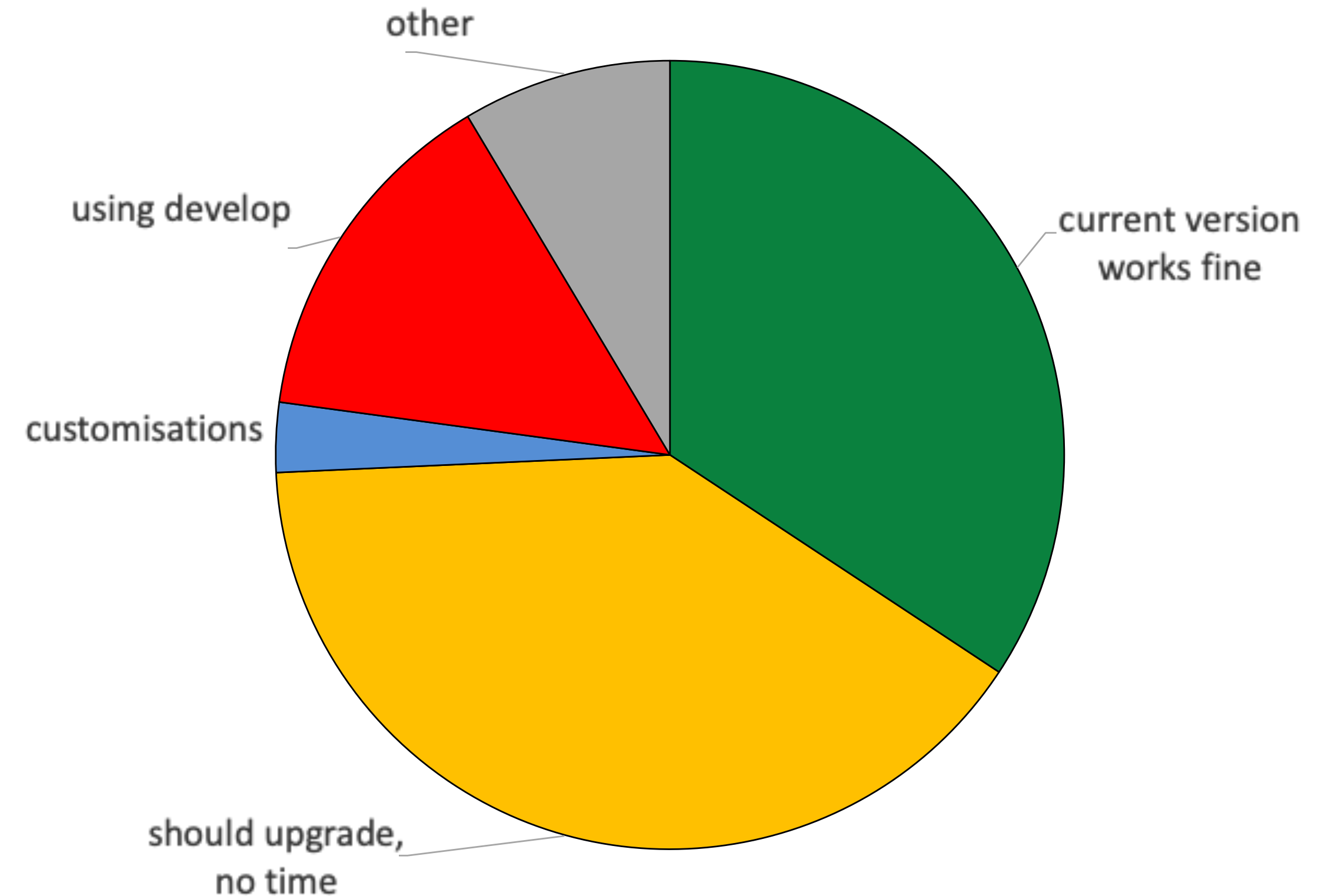
Which EasyBuild version do you use?



90% relies on EasyBuild releases,
80% is using EasyBuild 4.3.x,
small minority using develop (only maintainers?)

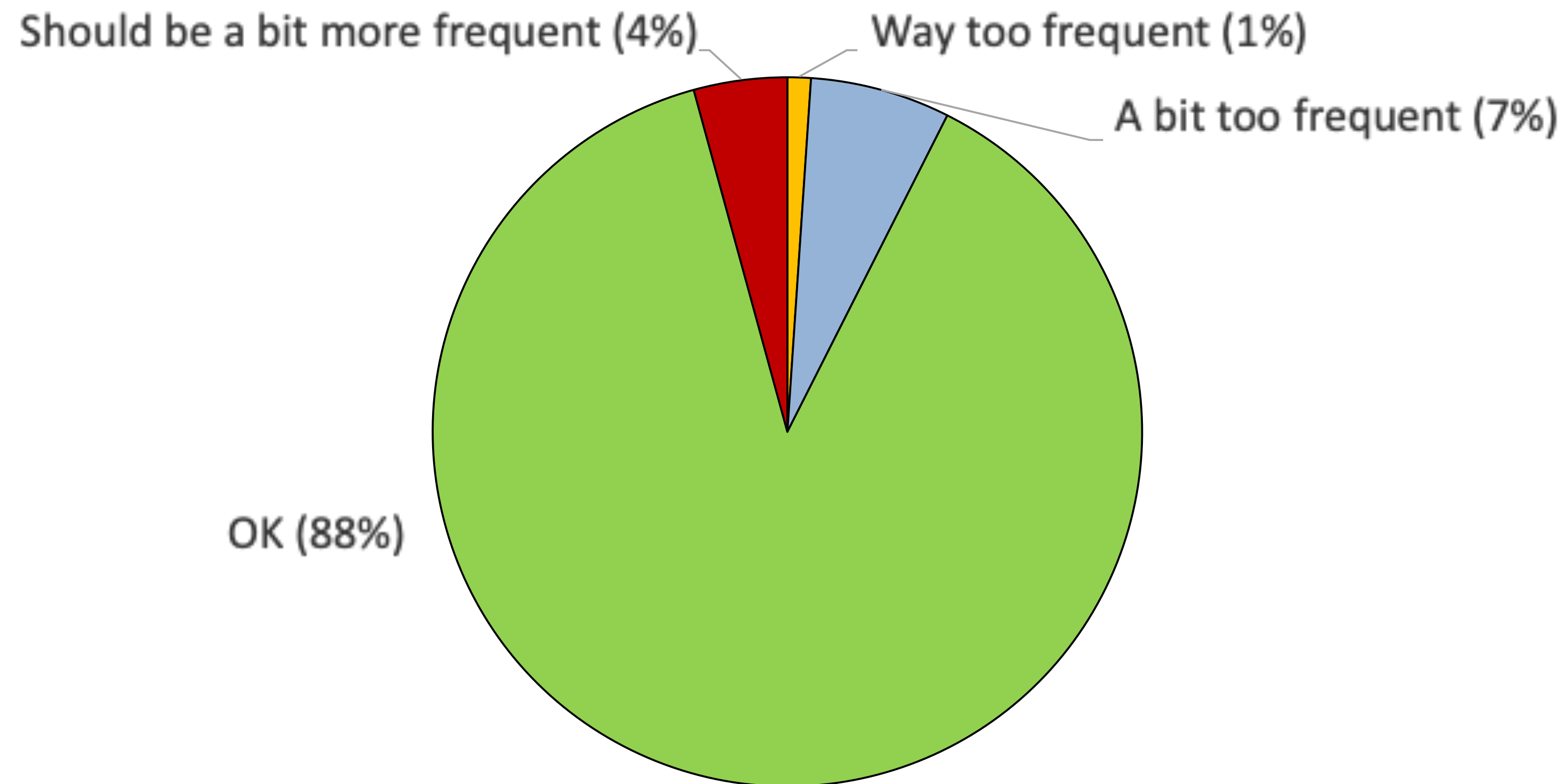
If you're not using the latest release, why not?

4 out of 10 mentioned something else than "using latest release":



overall very similar to previous survey

How do you like the frequency of EasyBuild releases?

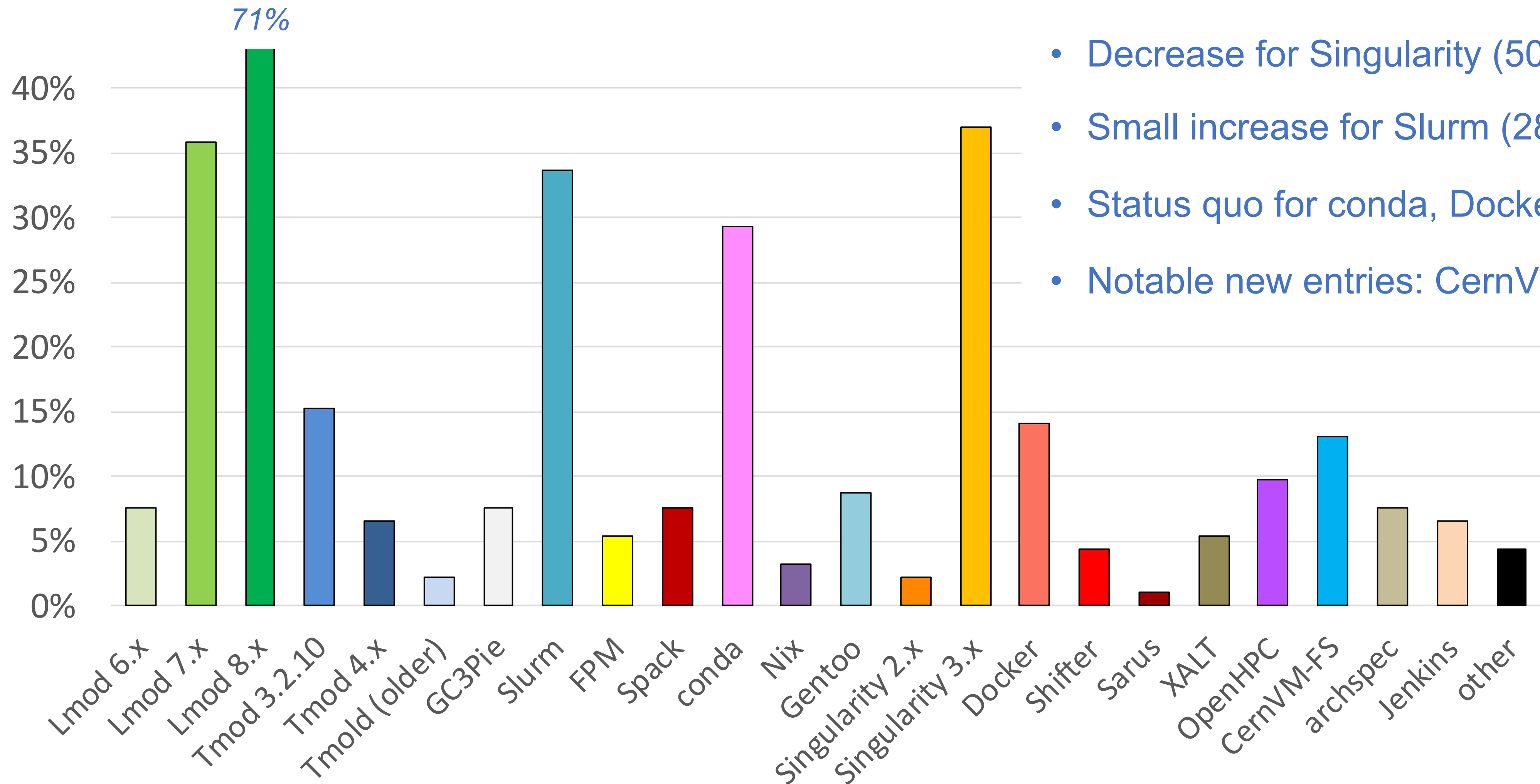


vast (and growing) majority is OK with current release frequency (75% ➡ 88%),
slightly more "too frequent" than "a bit more frequent"

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



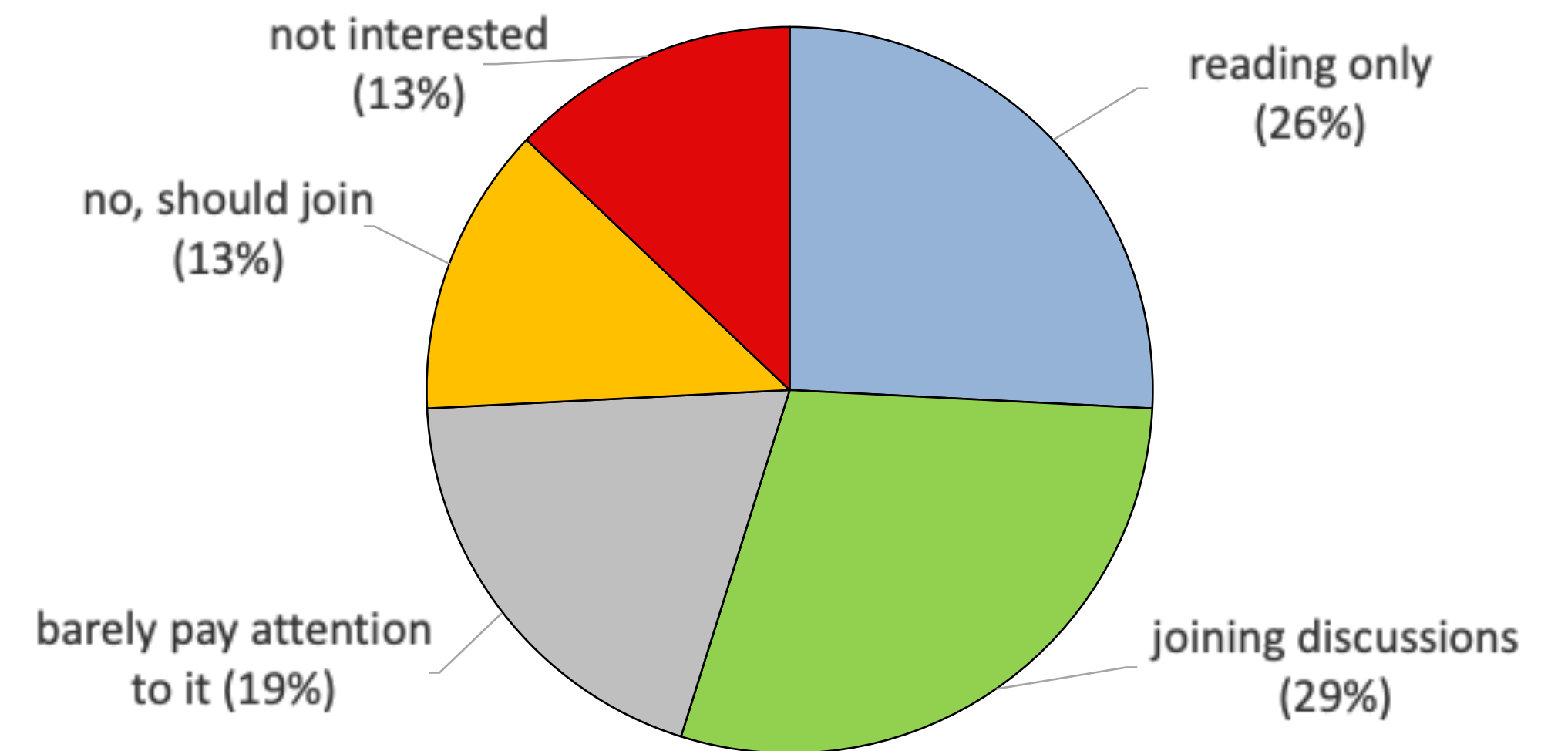
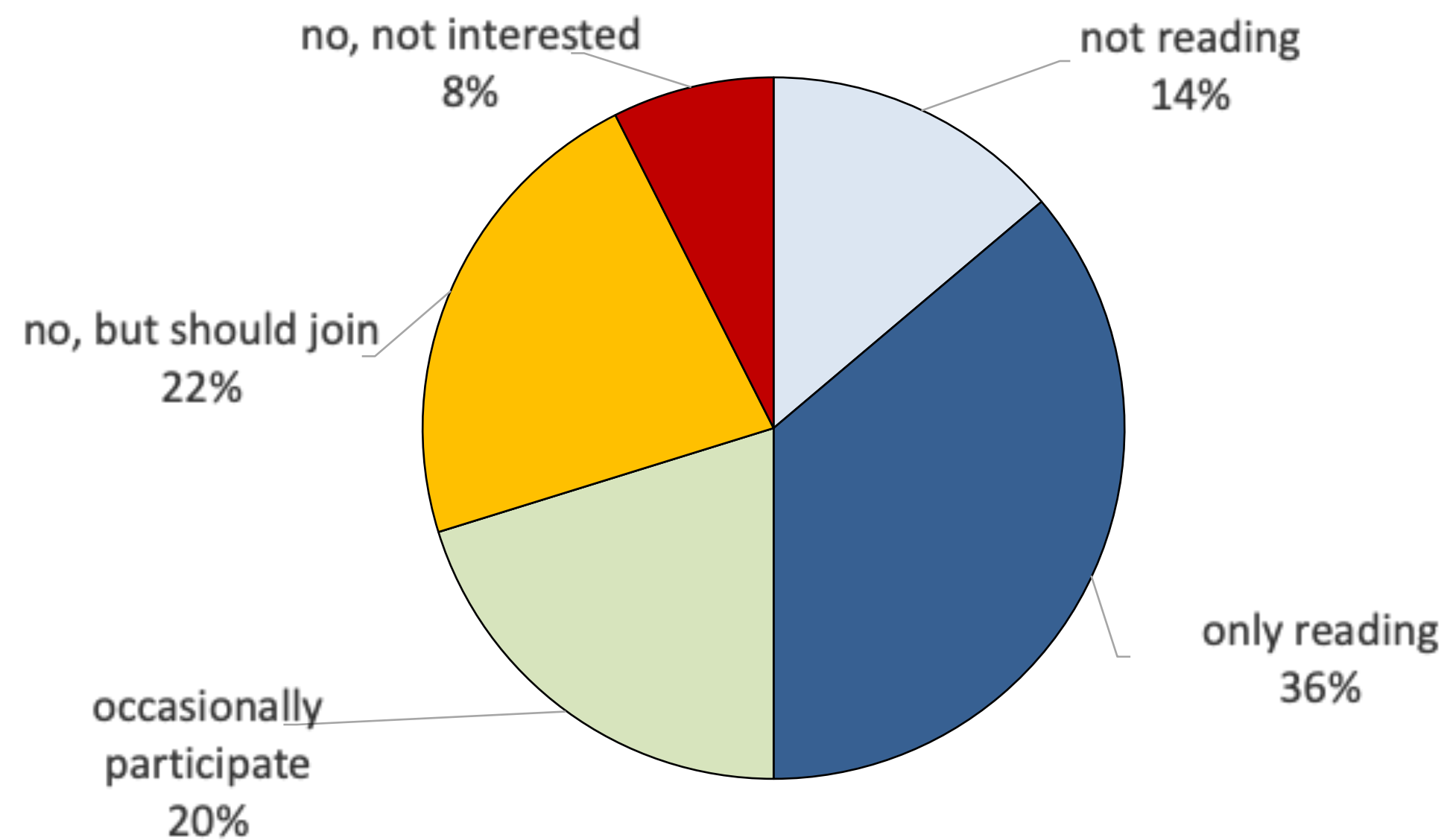
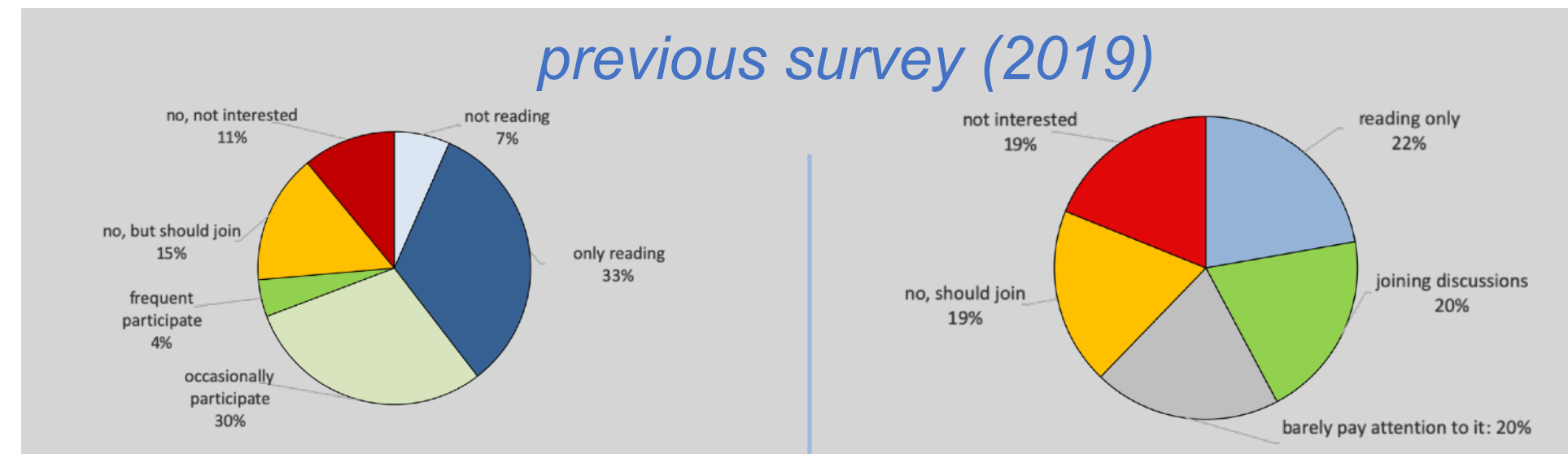
Do you use any other tools/projects in combination with EasyBuild?



- Vast majority uses Lmod as modules tool
- Decrease for Singularity (50% → 39%),
- Small increase for Slurm (28% → 33%)
- Status quo for conda, Docker, OpenHPC, Spack
- Notable new entries: CernVM-FS, archspec

Are you subscribed to the EasyBuild mailing list?

Do you use the EasyBuild IRC/Slack channel?

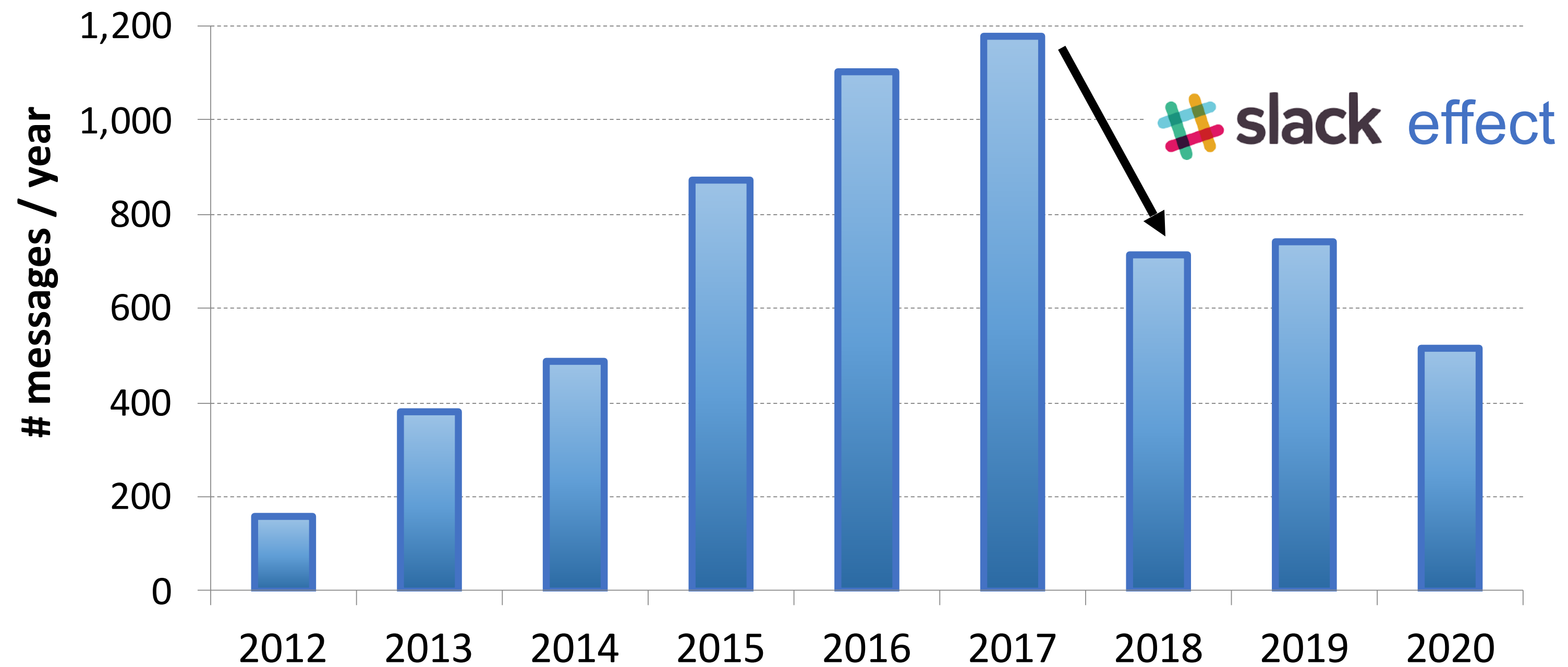


Interest in (active) participation in mailing list is decreasing,
interest/use of Slack is increasing.

A significant part of the EasyBuild community remains silent!

easybuild mailing list (traffic)

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

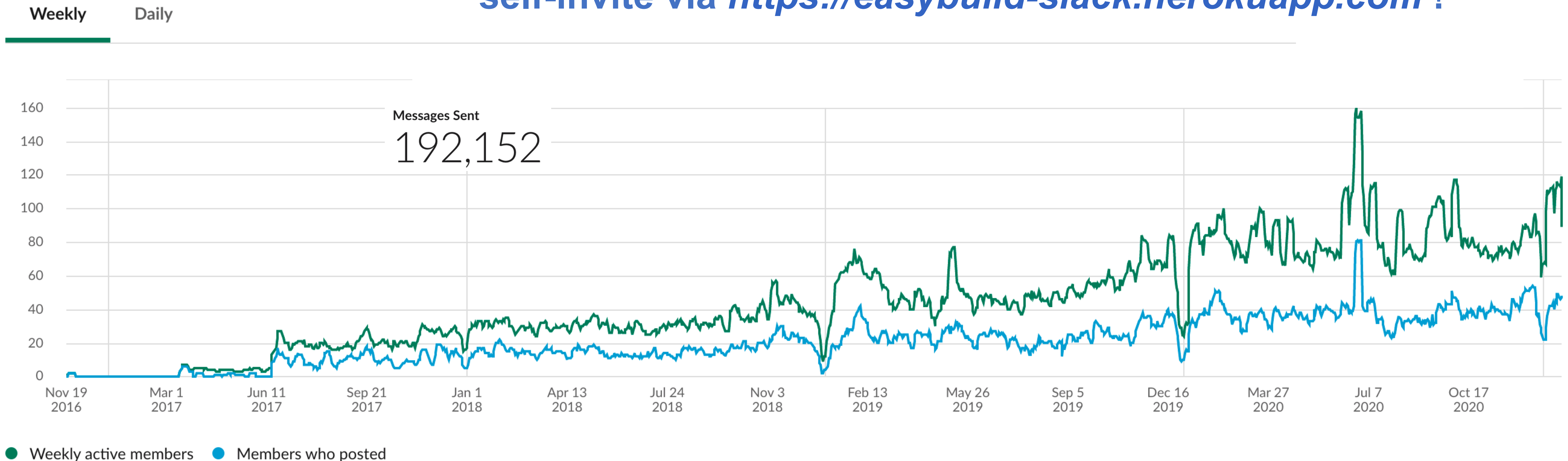


281 subscribers to EasyBuild mailing list currently
(subscribe via <https://lists.ugent.be/wws/subscribe/easybuild>)

easybuild # slack channel

- steady growth in activity since start of EasyBuild Slack in 2017
- 365 members on Slack channel (+70 in 2019, +160 in 2020!)
- Total messages sent almost *doubled* in 2020 (105k ➡ 192k)

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

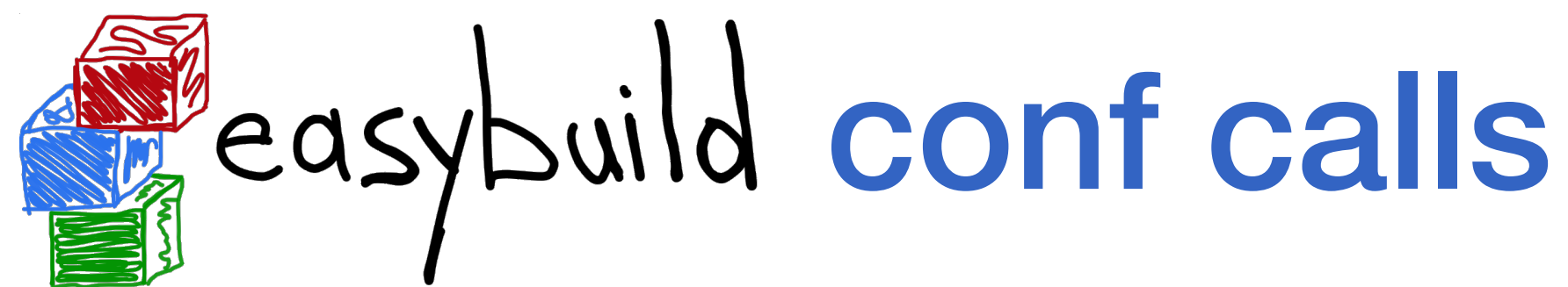




easybuild Tech Talks

<https://github.com/easybuilders/easybuild/wiki/EasyBuild-Tech-Talks>

- Idea emerged to invite speakers to do in-depth technical talks
- On topics relevant to the EasyBuild user community
- Sessions are open to anyone interested, streamed + recorded via YouTube
- So far:
 - **"The ABCs of OpenMPI"** by Jeff Squyres and Ralph Castian (3 parts!)
 - **"Yes! You Can Run Your Software on Arm"** by Chris Edsall
- Suggestions for topics are welcome!

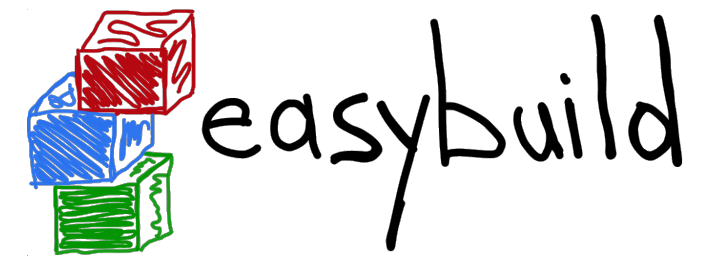


easybuild conf calls

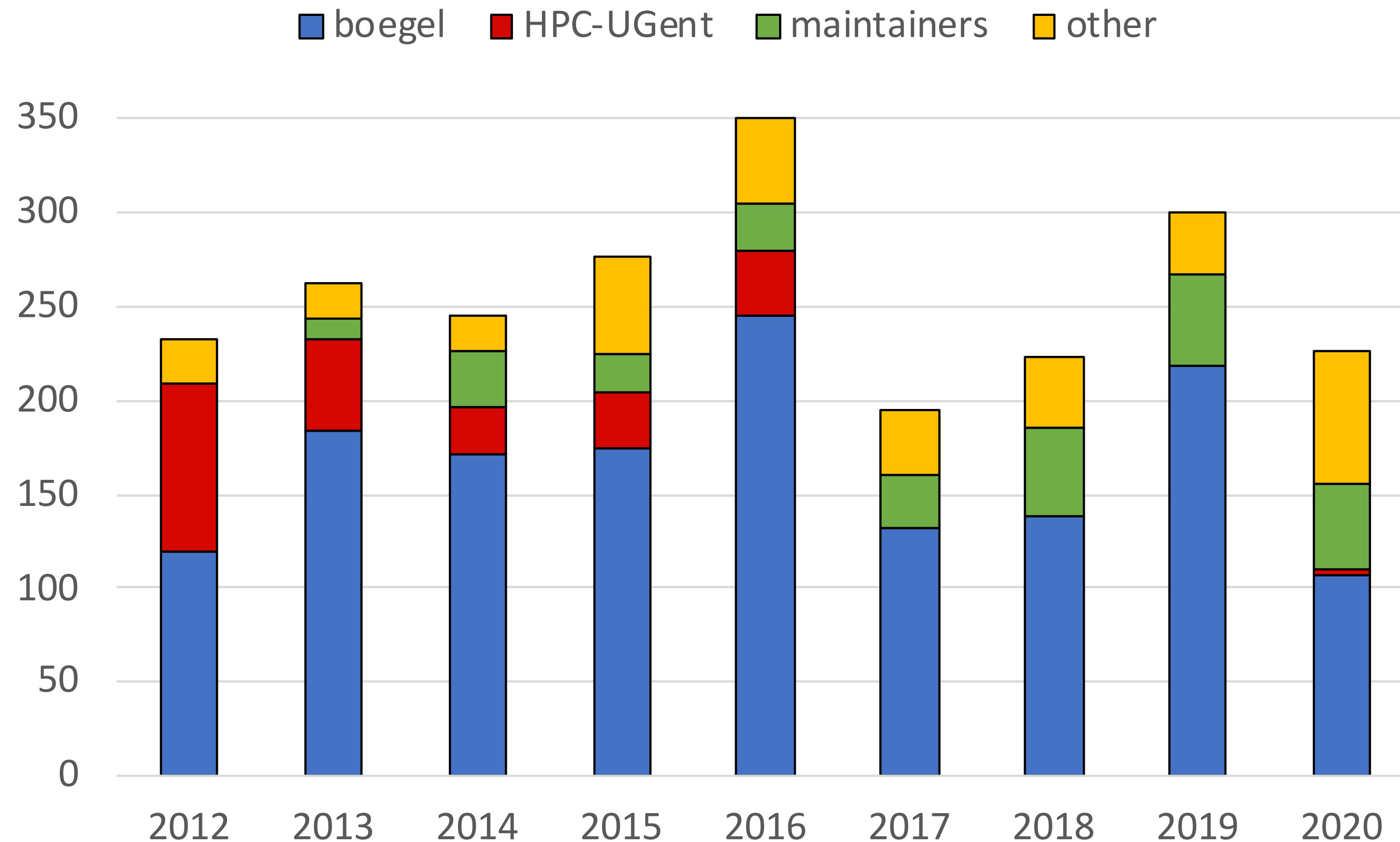
<https://github.com/easybuilders/easybuild/wiki/Conference-calls>

- EasyBuild conference calls via Zoom, every other week
- Open to anyone to join who is interested in EasyBuild
- Since Nov. 2013, barely any were skipped => **165** conf calls so far!
- Recently started switching between 10am CET and 5pm CET
(was only 5pm CET before)
- Usual topics: overview of recent developments, plans towards next release, Q&A
- Additional topics based on recent discussions or following suggestions
- Detailed notes are kept on EasyBuild wiki

Contributors, contributors, contributors...

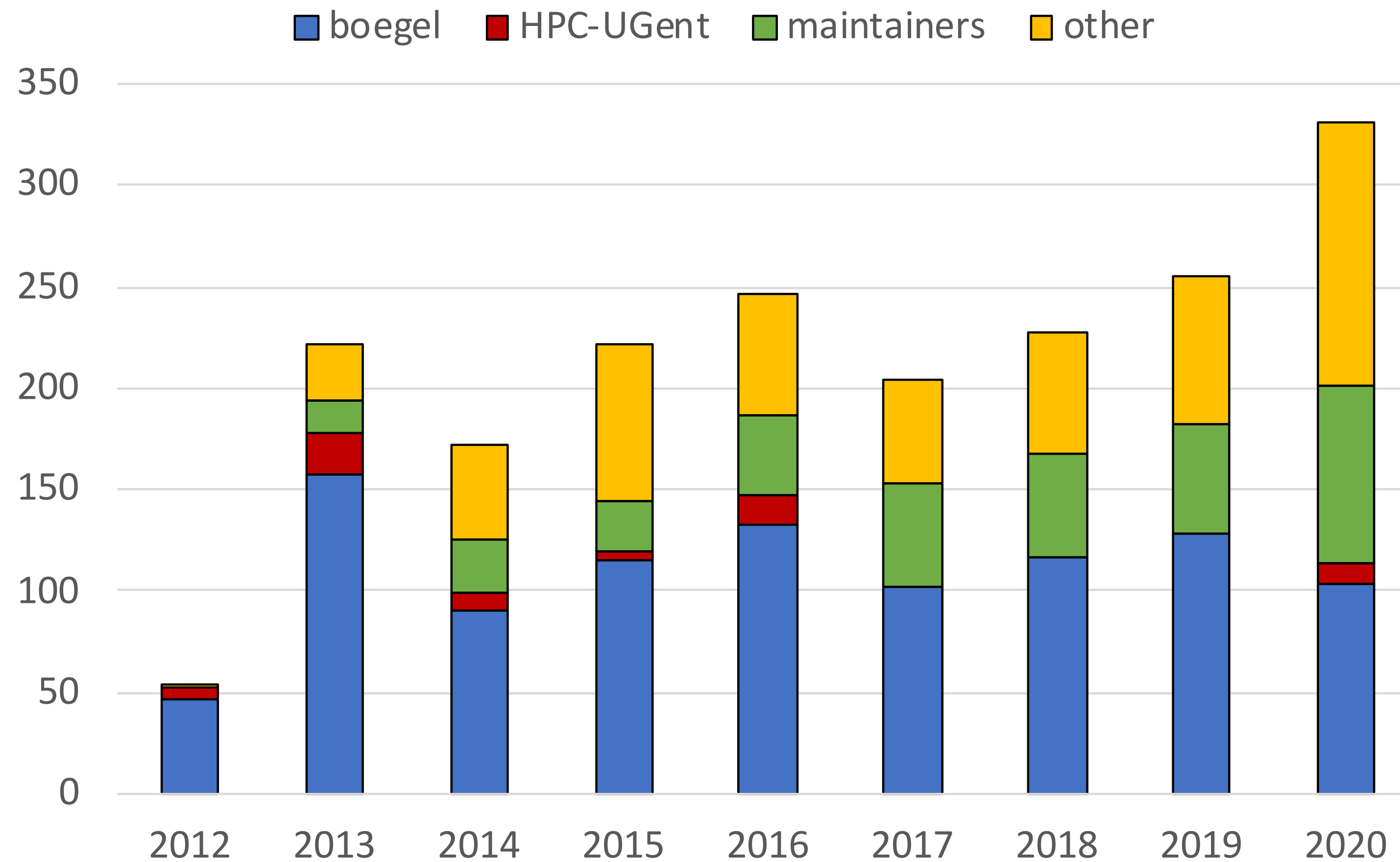


Overview of contributions (framework)



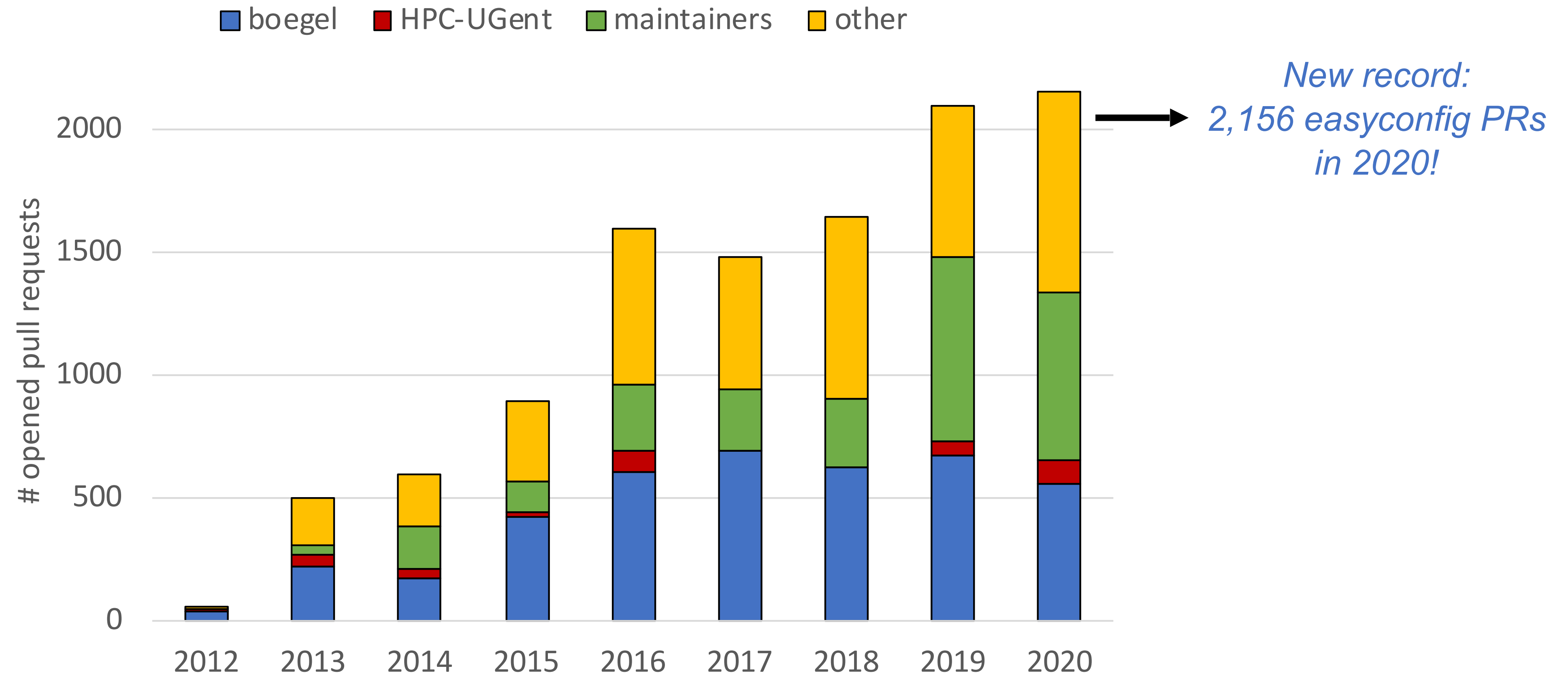
- > 50% of framework PRs in 2020 by contributors outside HPC-UGent
- Peak in PRs in 2019 is due to effort for porting EasyBuild to Python 3

Overview of contributions (easyblocks)



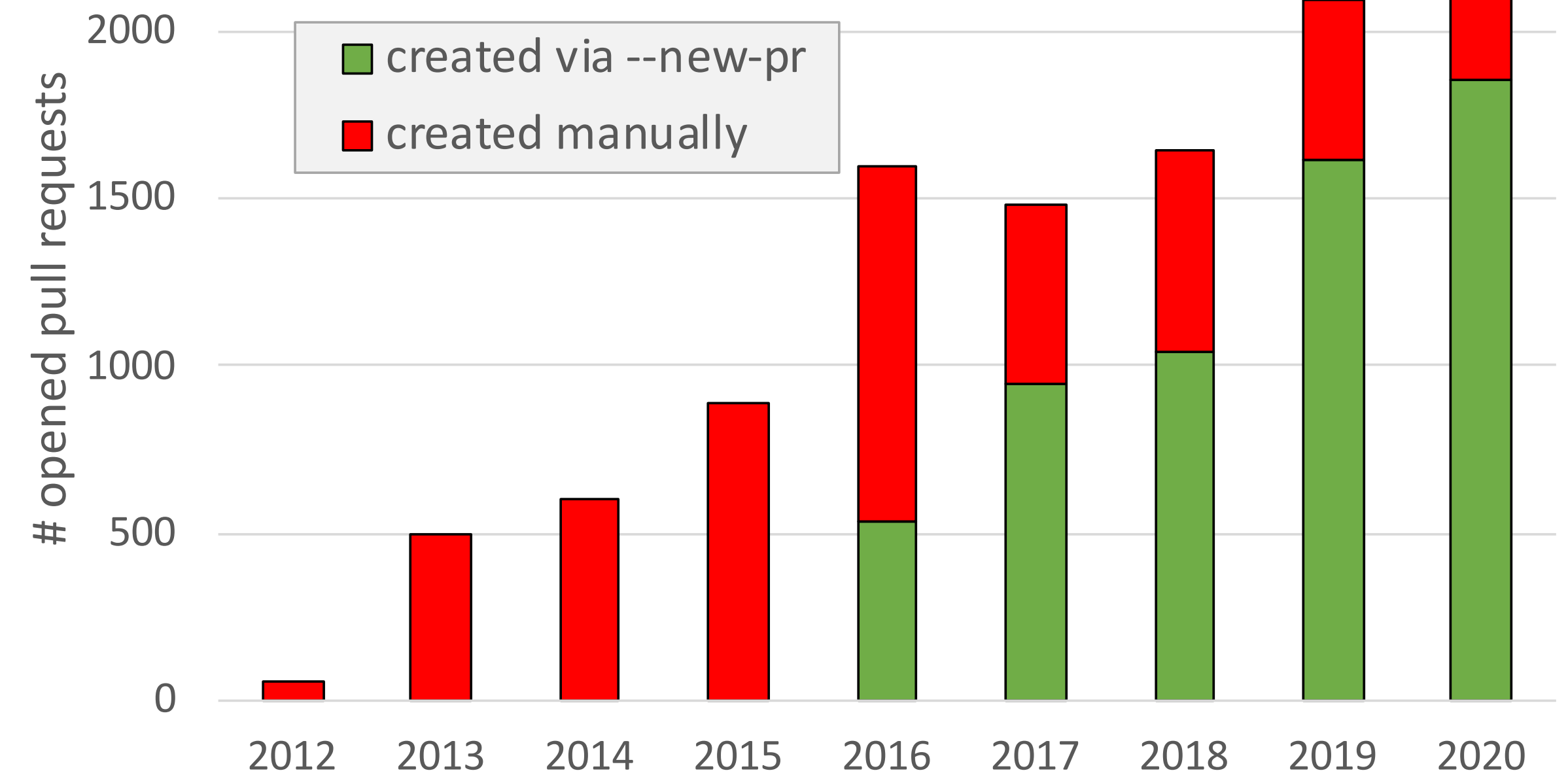
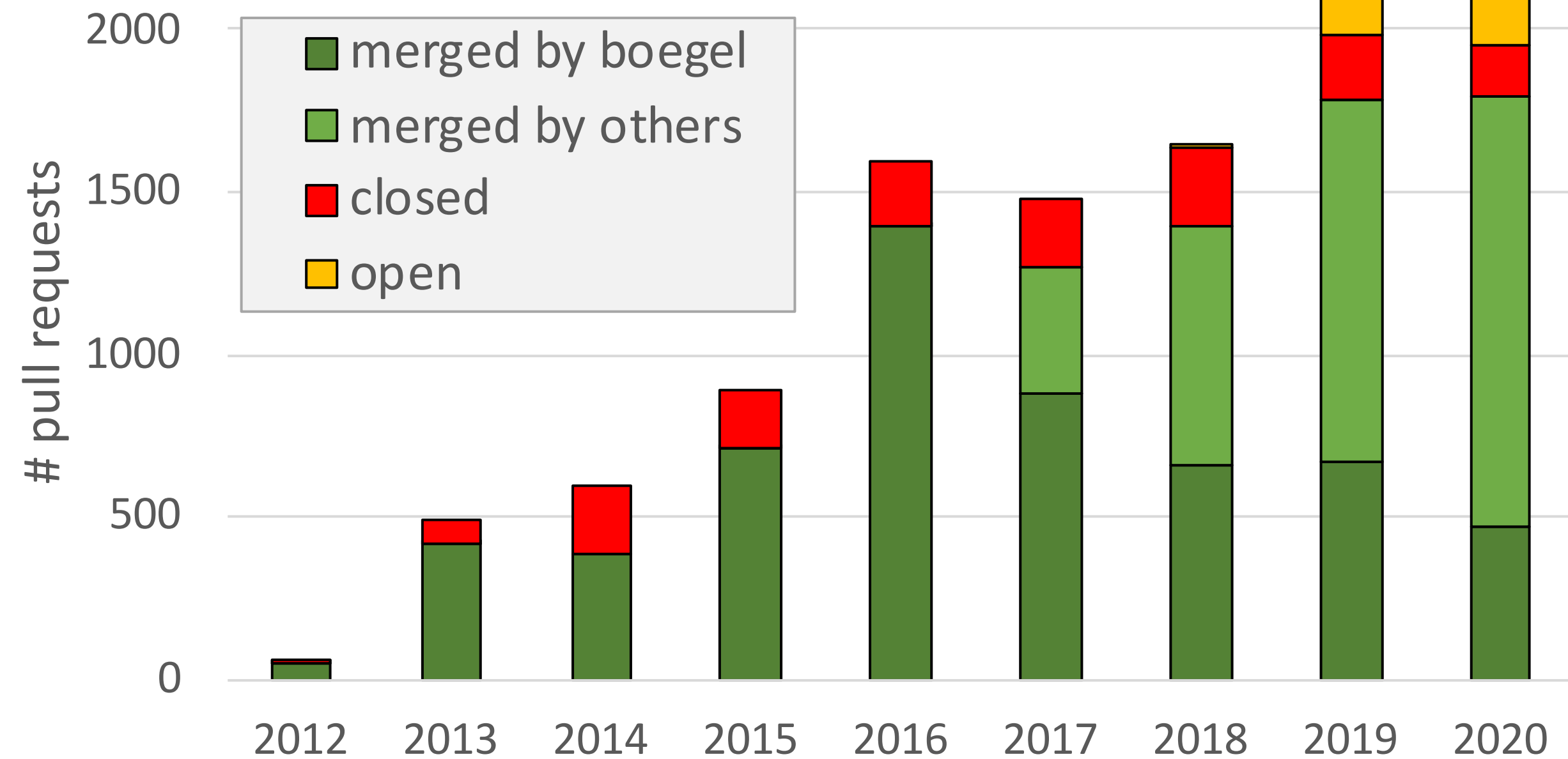
- Significant increase in contributions compared to previous years (+29% vs 2019)
- > 65% of PRs by outside contributors, 39% by people who are not an EasyBuild maintainer

Overview of contributions (easyconfigs)



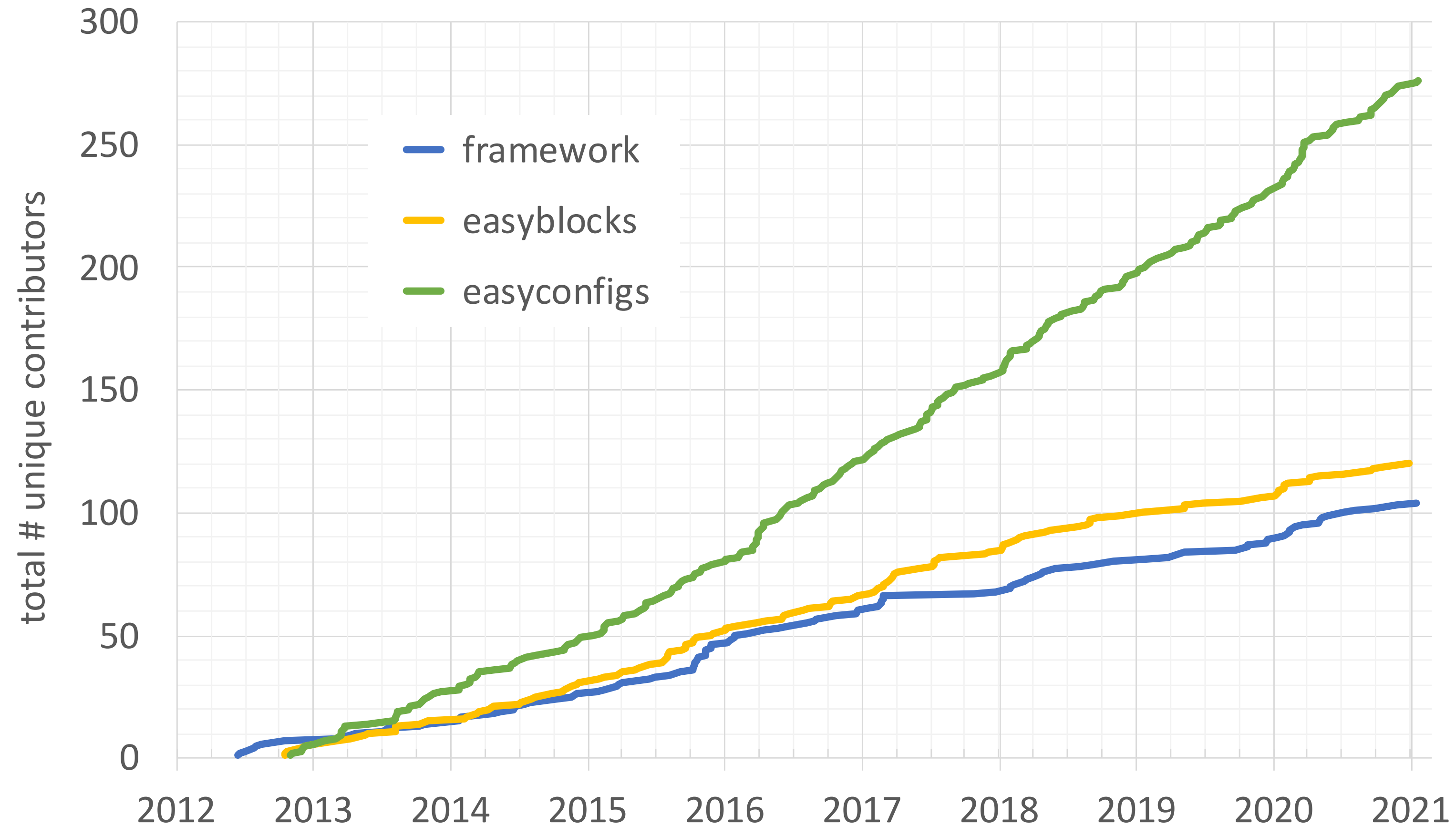
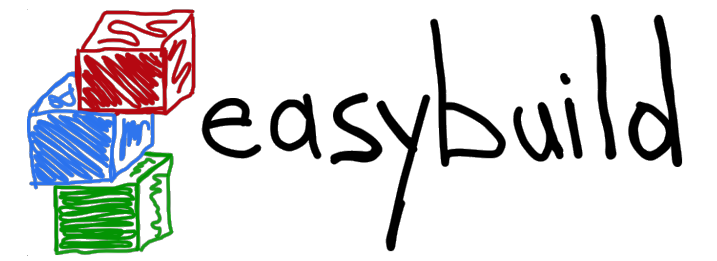
- explosive growth in easyconfig PRs in 2016 (partially thanks to `--new-pr`)
- stabilised in 2016-2018, significant increase in 2019-2020: > 2,000 PRs...
- growing ratio of PRs from contributors outside of HPC-UGent (53% ➡ 62% ➡ 65% ➡ 70%)

Overview of contributions (easyconfigs)



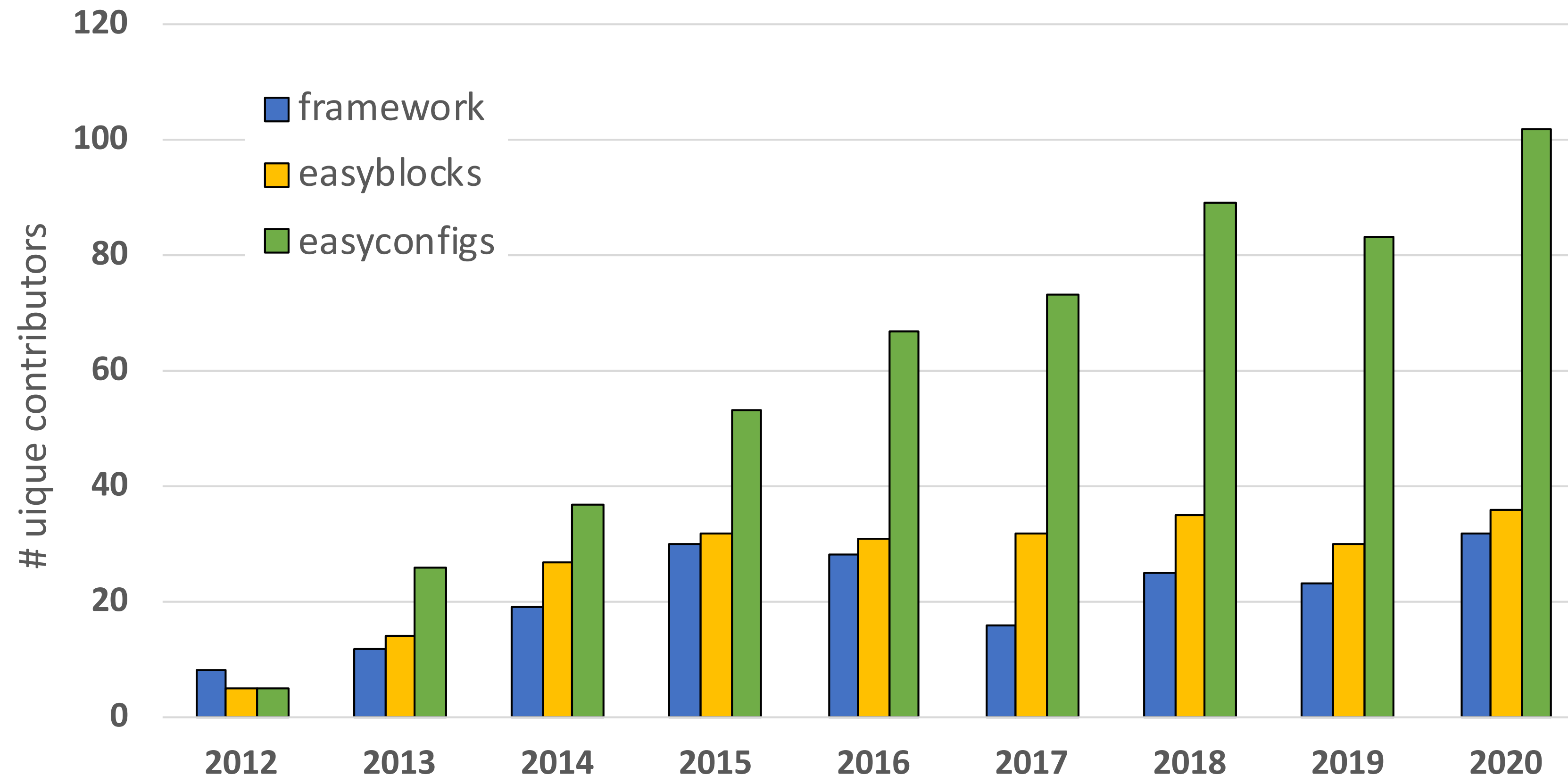
- growing number of easyconfig PRs merged by other maintainers (~74% in 2020)
- ratio of PRs opened with "eb --new-pr" is increasing even more (86% in 2020!)

Overview of unique contributors (total)



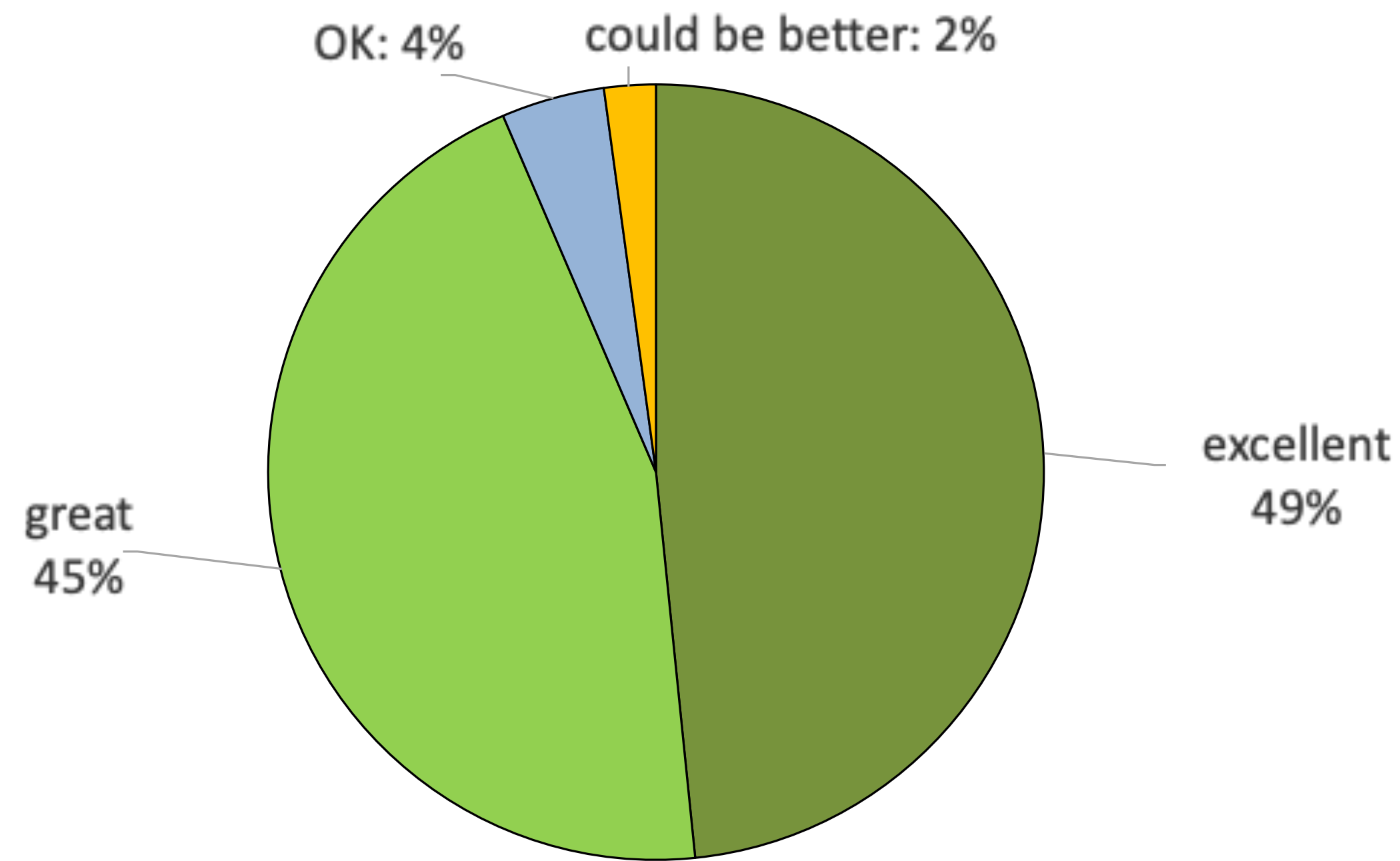
- Over 100 unique contributors to EasyBuild framework so far (+14 in 2020)
- continued growth in unique contributors for easyblocks (120, +14 in 2020)
- steadily growing number of unique contributors to easyconfigs (soon 300?)

Overview of unique contributors (per year)



- ~30 unique contributors per year to EasyBuild framework
- 30-35 unique contributors per year to easyblocks, fairly stable since 2015
- 80-100 unique contributors per year for easyconfigs (102 in 2020!)

How would you rate the overall quality of EasyBuild?



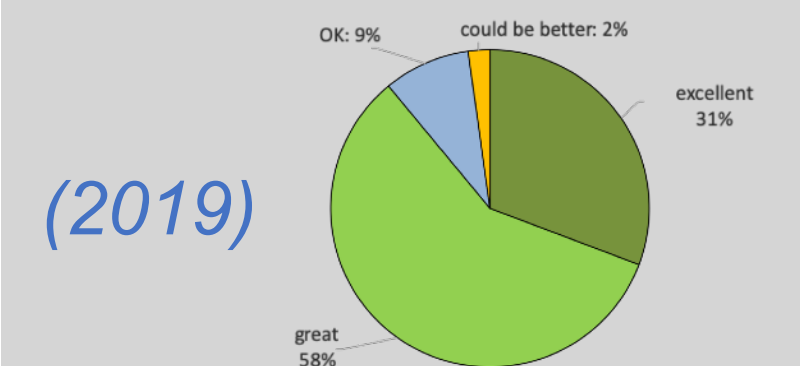
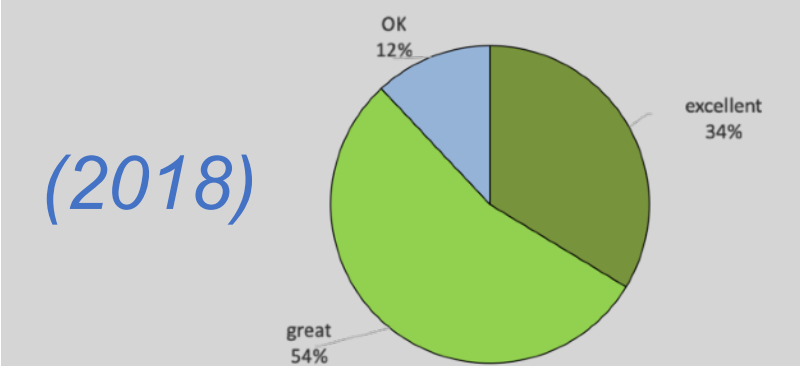
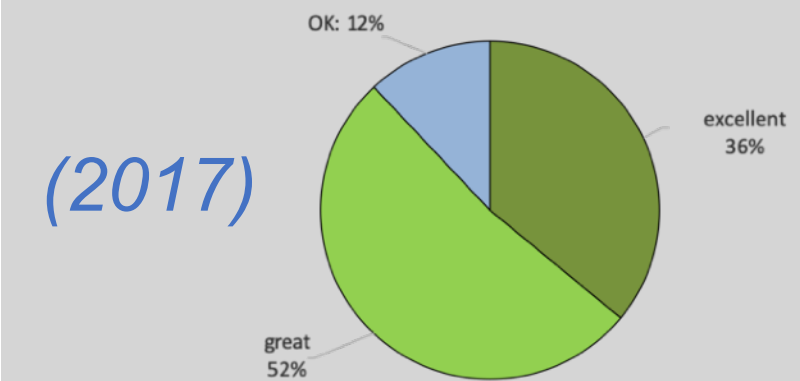
98% positive!

Significant increase in "excellent" (31% ➡ 49%)

"Great" + "excellent" slightly increasing... (91% ➡ 94%)

(+ nobody answered "pretty bad")

previous surveys



Goals for 2021



- Better error reporting: isolate actual problem if shell command failed, ...
- Make transitioning of existing easyconfigs to other toolchains (even) easier
- Improved EasyBuild regression test (+ public dashboard?)
 - Using Re²Frame
 - Using container images to control test environment (OS, installed packages, ...)
 - On variety of processor architectures (different EasyBuild sites, AWS?)
- Rework documentation with `mkdocs` (cfr. EasyBuild tutorial)

Challenges in 2021



- Insatiable hunger of scientists for getting more (new) software packages installed...
- **Increasing diversity in hardware (CPU/GPU) and software (toolchains)**
- Keeping up with incoming contributions/pull requests
 - As a contributor: please use GitHub integration features (`--new-pr` & co)
 - Bot(s) should take care of more stuff...
 - *Automatic* test reports once CI tests pass + approved review by maintainer
 - For variety of CPU architectures and OS configurations (using containers)

Increasing diversity in hardware & software



- Hardware:
 - CPUs: Intel, AMD, Arm (different vendors...), POWER, perhaps also RISC-V (cfr. EPI)
 - GPUs: NVIDIA, AMD Radeon, Intel Xe
- Software: Intel oneAPI, AOCC (AMD CPUs), ROCm (AMD GPUs), NVHPC, Clang, ...
- We need to rethink approach to (common) toolchains?
 - Merging foss + fosscuda: make CUDA a "dynamic" dependency for OpenMPI?
 - Additional Clang-based common toolchain?
- Considering to switch away from OpenBLAS to BLIS in foss (/2021a) toolchain
 - Initial performance results on AMD Rome look very promising!
 - Further testing and benchmarking needed





EasyBuild + LUMI, the Queen of the North

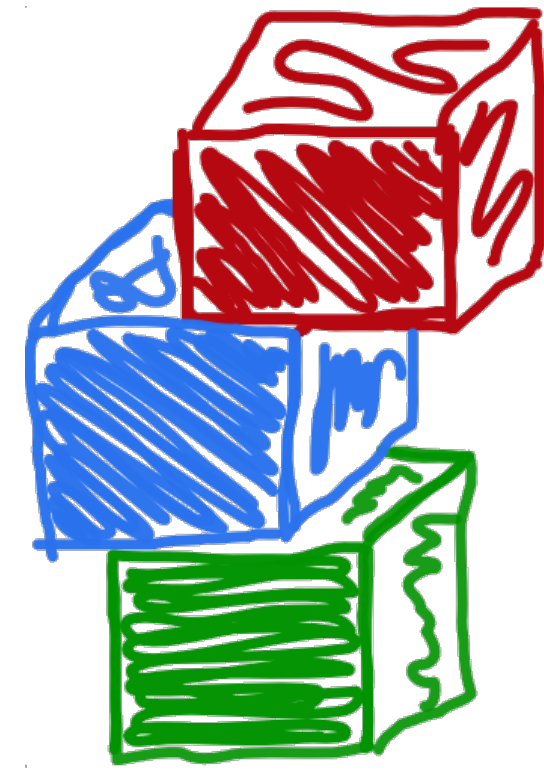


L U M I

www.lumi-supercomputer.eu

- The LUMI User Support Team (LUST) has been evaluating software installations tools...
- Two major contenders:  easybuild and  **Spack**
- Final decision was made public **today**...
- **EasyBuild has been selected as the (primary) build tool for LUMI !**
- AMD CPUs + GPUs, cfr. <https://www.lumi-supercomputer.eu/may-we-introduce-lumi>
- *"We are looking forward to start collaborating with the EasyBuild community!"*





easybuild

Questions?

[@boegel \(GitHub\)](#)

[@kehoste \(Twitter\)](#)

kenneth.hoste@ugent.be

easybuild@lists.ugent.be

<https://easybuild.io>

<https://docs.easybuild.io>