



RDKit UGM 2020

Wherever you are

https://github.com/rdkit/UGM_2020

 #RDKitUGM2020

Introductions

Who am I?

Greg Landrum (T5 Informatics GmbH, KNIME AG)

Who organized this?

Who are you?

Please join the discord chat: <https://discord.gg/xxjzUW3>

Thanks to our friends at KNIME for, once again, supporting the RDKit UGM

Introductions

Who am I?

Who organized this?

Floriane Montanari, Djork-Arné Clevert (Bayer)
Christiane Kallfass (KNIME)

Who are you?

Please join the discord chat: <https://discord.gg/xxjzUW3>

Thanks to our friends at KNIME for, once again, supporting the RDKit UGM

Wow!

eventbrite

View

• On Sale ▾

More ▾

GL ▾

Virtual 9th RDKit UGM 2020

Tue, Oct 6, 2020 9:00 AM

[Switch Event](#)



Basic Info



Details



3 Online Event Page

Tickets

Admission

Add-ons

Holds

Settings

Create Ticket



General Admission

• On Sale • Ends today at 4:00 PM

584 / 650

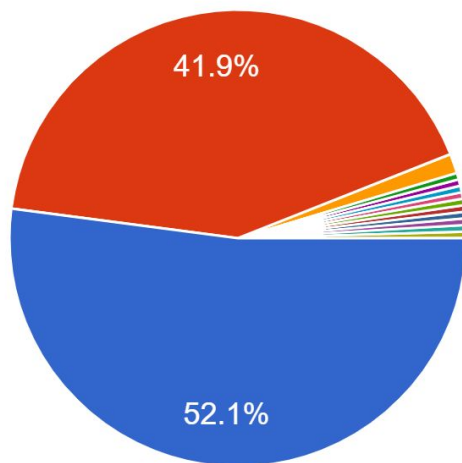
Free



Who are you?

Where do you work?

217 responses

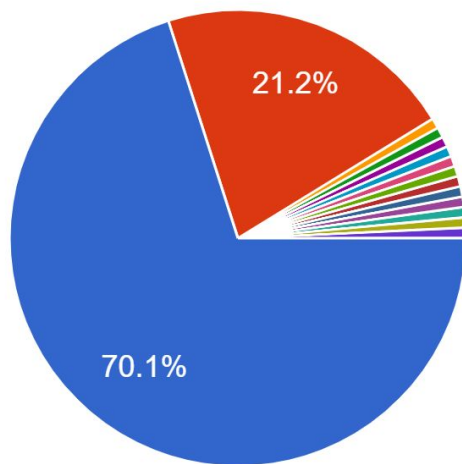


- Industry
- Academia
- both
- Research Institute (non academia)
- US Food and Drug Administration
- Non-profit research organization
- Non profit organization
- SME

Who are you?

If you're in industry, what kind?

137 responses

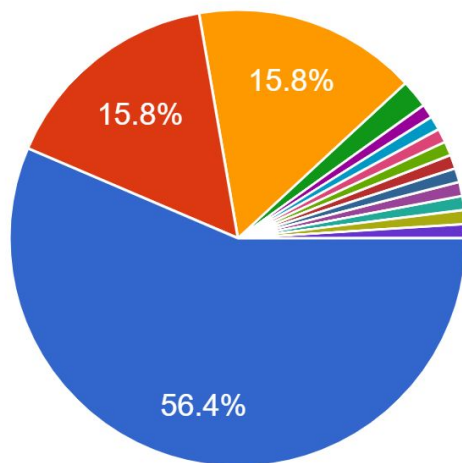


- Pharma/biotech
- Software
- Both: software-driven biotech
- Chemicals
- Consulting
- Crop Science
- Chemistry
- Agrochemistry

Who are you?

If you're in academia, what do you do?

101 responses

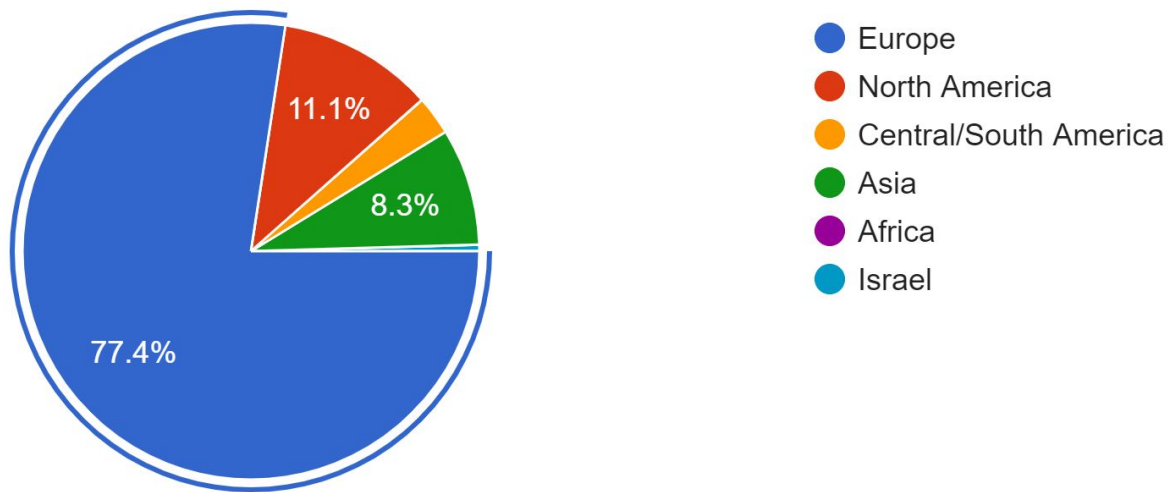


- Student
- Postdoc
- Faculty
- Staff Scientist
- Research Technician
- Just graduated
- Professor/Researcher
- Software engineer

Who are you?

Where are you?

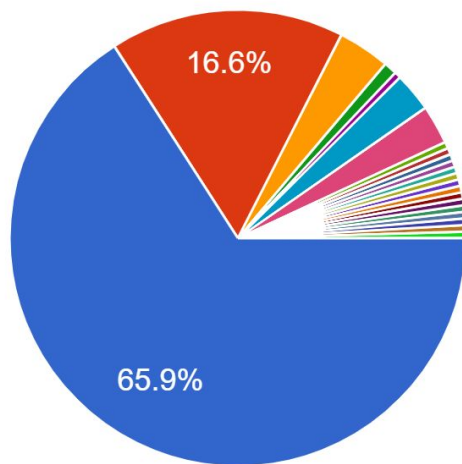
217 responses



Who are you?

How many RDKit UGMs have you attended?

217 responses

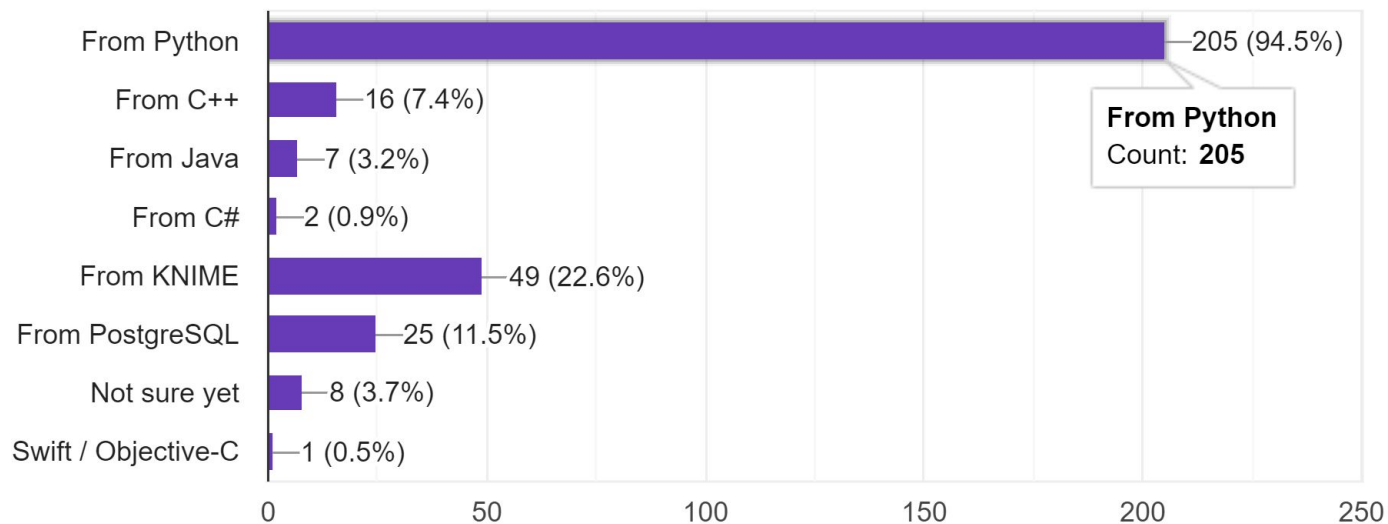


- This is my first one!
- Three or more
- Five or more
- All of them except one
- All of them
- 1
- 2
- Only one, RDKit UGM 2019

Who are you?

How are you using the RDKit?

217 responses



Who are you?

Have you contributed to the RDKit and/or the community?

82 responses



“Frequent Hitters”

- Nine UGMs:
 - Roger S
 - Grégori G
- Eight UGMs:
 - Sereina R
 - Paul C

If you should be on here and aren't: please let me know!

Aside: Past meetings/organizers

1. London (ICR): Nathan Brown
2. Hinxton (EMBL): George Papadatos
3. Darmstadt (Merck): Paul Czodrowski
4. Zurich (ETH): Sereina Riniker
5. Basel (Novartis/Roche): Nadine Schneider, Nik Stiefl, Christian Kramer
6. Berlin (Charité Berlin/FU Berlin): Andrea Volkamer, Gerhard Wolber
7. Cambridge (Cambridge University): Andreas Bender
8. Hamburg (Uni Hamburg): Emanuel Ehmki

About the meeting

- Friendly
- Diverse
 - People
 - Topics
 - Degrees of technicality
- By/for the community
 - Organized and run by volunteers

Please join the discord chat: <https://discord.gg/xxjzUW3>

Thanks to our friends at KNIME for, once again, supporting the RDKit UGM

A quick aside about toolkits

The RDKit is a toolkit

- It provides a bunch of tools that you can combine to build interesting and useful things
- It doesn't provide a UI, a set of web services, or a series of scripts that you can run¹; those are generally found in other projects. Or you build them yourself. :-)

¹ The exception to this is some of the stuff that you will find in Contrib: those are interesting things that other people have built.

Next topic: community

Community

The heart of any
successful open-source
project



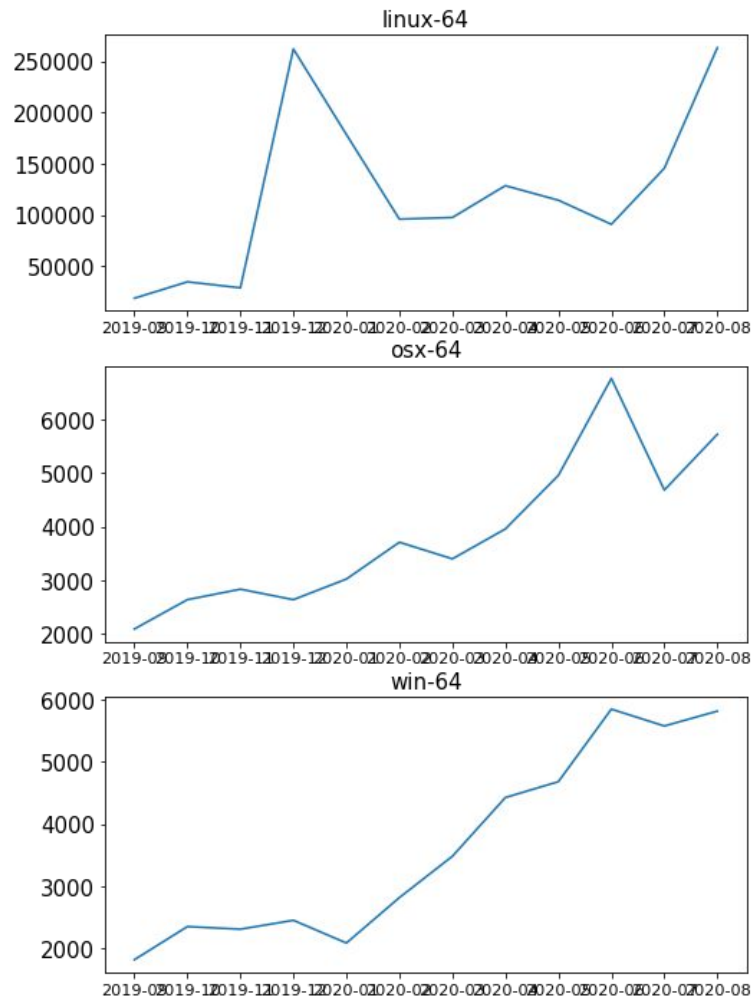
Community: the heart of any successful open-source project

- Mailing lists: >1000 messages to rdkit-discuss from 2019.09.28 - 2020.09.28
- Google scholar: >850 hits for "rdkit" in 2019
- Searching github for `from rdkit`
`import Chem` returns >12000 code results in Python files. These are contained in >500 unique repo names (i.e. not RDKit forks).
- Each of the last six UGMs at capacity with 40-100+ attendees
- Seven Google Summer of Code projects over the last four years

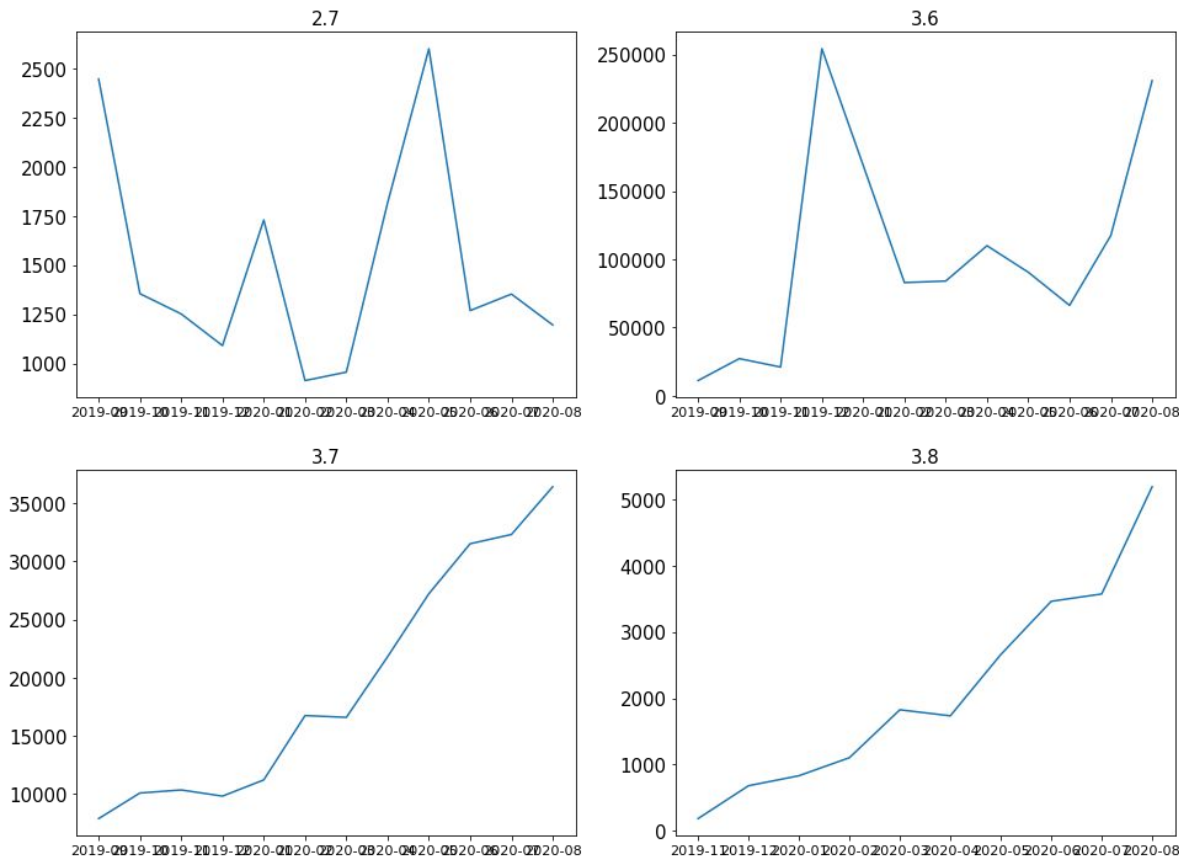


Usage: Conda install counts (by operating system)

Last 12 months
Data collected using the
`condastats` package



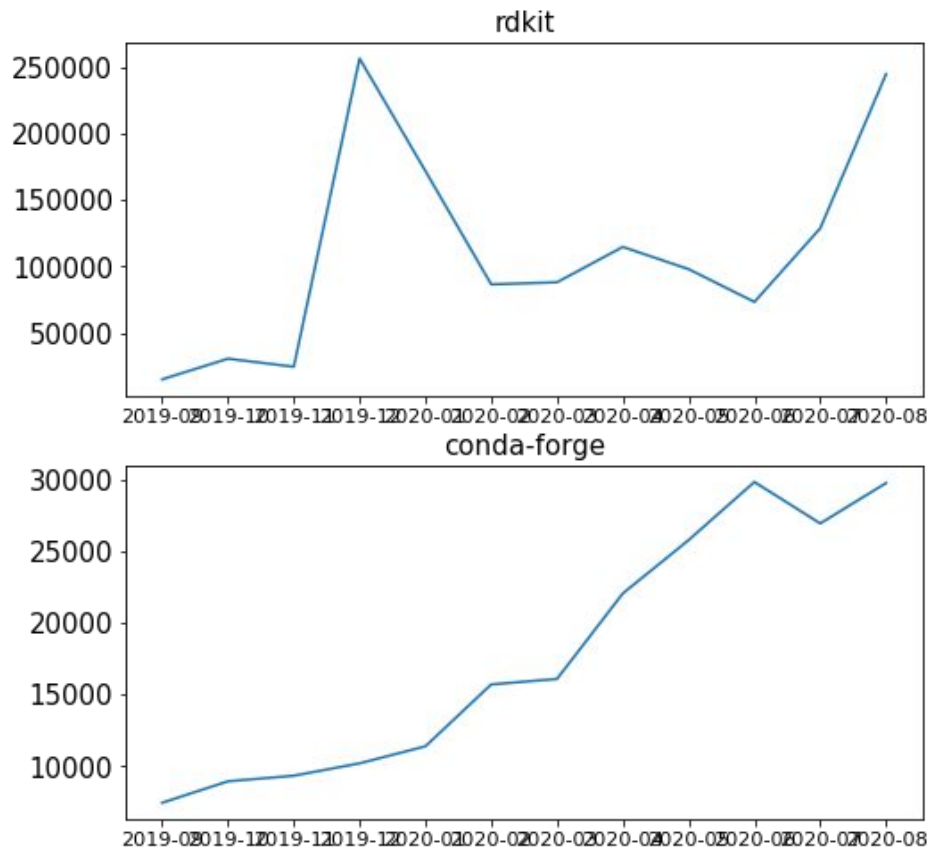
Usage: Conda install counts (by python version)



Last 12 months
Data collected using the
condastats package



Usage: Conda install counts (by source)



Last 12 months
Data collected using the
condastats package



2019: Usage in other open-source projects

- stk (docs, paper) - a Python library for building, manipulating, analyzing and automatic design of molecules.
- OpenFF - Open source approach for better force fields
- gpusimilarity - GPU implementation of fingerprint similarity searching
- Samson Connect - Software for adaptive modeling and simulation of nanosystems
- mol_frame - Chemical Structure Handling for Dask and Pandas DataFrames
- mmpdb 2.0 - matched molecular pair database generation and analysis
- CheTo - Chemical topic modeling
- OCEAN - web-tool for target-prediction of chemical structures which uses ChEMBL as datasource
- Coot - software for macromolecular model building, model completion and validation
- DeepChem - deep learning toolkit for drug discovery
- sdf_viewer.py - an interactive SDF viewer
- sdf2ppt - Reads an SDF file and displays molecules as image grid in powerpoint/openoffice presentation.
- chemfp
- PYPL - Simple cartridge that lets you call Python scripts from Oracle PL/SQL.
- WONKA - Tool for analysis and interrogation of protein-ligand crystal structures
- OOMMPAA - Tool for directed synthesis and data analysis based on protein-ligand crystal structures
- RRDKit - RDKit integration for R
- chemicalite - SQLite integration for the RDKit
- django-rdkit - Django integration for the RDKit
- ... more ...



Usage in commercial tools

- Cresset Software
- Dalke Scientific Software
- NextMove Software
- Schrödinger
- SCM
- Wolfram Research

Disclaimer: this info is from public statements made by people from those companies.
I almost certainly have forgotten someone



Usage in online tools/resources

- ChEMBL
- ZINC
- Google Patents
- PDBe
- Enamine
- TeachOpenCADD

Disclaimer: this info is from public statements made by people associated with those projects. I almost certainly have forgotten someone



Some community experiments

Trying to find a more interactive place to ask/answer RDKit questions and share news and expertise.

Two active experiments:

- Discord (more of a chat system):
<https://discord.gg/F9BF5Tn>
- Discourse (more of a forum):
<https://rdkit.discourse.group/>

We will try both for a while and keep one if interest is there



Support

- Web searches
 - Mailing list
 - Github
-
- Commercial support



Sustainability: the bus problem



https://commons.wikimedia.org/wiki/File:Postauto_susten.jpg



Sustainability: the bus problem

RDKit maintainers:

- Greg
- Brian Kelley
- Ricardo Rodriguez
- Paolo Tosco



Github commits in the last year

Sep 28, 2019 – Sep 27, 2020

Contributions: Commits

Contributions to master, excluding merge commits



greglandrum, daenuprobst, hjuinj, maksbotan, ricrogz, hsiaoyi0504, e-kwsm, d-b-w, jones-gareth, fdiblen, richardjgowers, sirbiscuit, lukasturcani, rachelnwalker, j-wags, autodataming, i-tub, pschmidtke, rvianello, shrey183, daverona, skearnes, ElianeBriand, zhentg, DavidACosgrove, bp-kelley, lummyk, jvansan, vfscalfani, wvandertoorn, manangoel99, yellowBirdy, yishutu, mwojcikowski, chriswzou, cornett, intrigus-lgtm, ptosco, thegodone, iammosespaulr, mcs07, UnixJunkie, chmnk

That's 43 different people

Contributions to github issue tracker in the last year

ErikCVik harsh6646 thegodone YuanyueLi xhj-down smoors b-mahjour karolbadowski danielfflaneza yiino kexul Dekken jcmminerlanl yuyang-0825 LivC182 arjd e-kwsm alongd vsomnath rogerwq darkcircle zineb1202 JLVarjo gmseabra kjelljorner andrewtarzia Darkless012 rmmrg tarungog wangxr0526 mc-robinson giallu ljn917 xavierholt qq87258580 SeanXiaoyuSun drmatthewclark tonka-balonka johnmay Mario-Liu lucianlschan mbanck nanhuayu frakruger rvianello ryszard314159 shayakhmetov jhjensen2 UnixJunkie rajathsalegame pkrezel ZZP12 xjalencas AustinApple hsiaoyi0504 maksbotan msteijaert wopozka stewu5 Andy-Wilkinson slochower xiaoruiDong reinaw1012 BillLawrence111 seal9 Flora-Feng0532 autodataming yy692 Ljohn89 PjFlan Npontillo DavidACosgrove cgh2797 ethan0879 chmnk NKUCodingCat i-tub ano302 ChongLu121 btrx-kheyer jannisborn lpravda Et9797 sroughley asfalcao mti193 ValeryPolyakov IgnacioJPickering Plancalkuele greglandrum MherMatevosyan jmechua-itsrdc hl2500 RuikangSun ZacharyKaplan ericmjonas khoivan88 spadavec yurivict IchiruTake ericmjl manangoel99 kangway roccomoretti pcko1 sdvillal jungsdao bokertof koyurion cdvonbargen andt88 matteoferla paconius Laksh1997 mkossner kheyer YuYaoYang2333 elfring xiaohongniua gfrogat IvanChernyshov bp-kelley XiangyanSJTU shenkin prerpd mwojcikowski adalke AustinT mitrandyrshrey183 ziedhosni jasondbiggs sbhakat proteneer mist714 marcusstamm d-b-w rd0102 danpol josevlibera2010 olgOk e-mayo soerendip genuinekale shirishag75 DrrDom i-shah jminuse pyeguy TheChymera david-turnbull yipy0005 markf94 s-ansa lpfSTX zhandn ricrogz jose-mr ptosco

That's 159 different people



Maintenance work in the last year

We started tracking maintenance/cleanup work with the 2019.09 release.

For the 2020.03 and 2020.09 releases, there have been 30 “cleanup” issues/PRs merged:

Greg Landrum 16
Eisuke Kawashima 6
Paolo Tosco 3
Ric 2
Moses Paul R 2
Riccardo Vianello 1
Dan N 1
Brian Kelley 1



Roadmap

Future work tends to be determined by what's needed for active projects or requests that come out of the community. So there's not much of a roadmap.



Still, some parts of the way forward are pretty obvious...

Making sure all the pieces required
to build a good compound
registration system are there

Making sure all the pieces required
to build a good corporate chemical
database are there

Performance improvements

Ongoing refactoring and code
cleanup



Ok, enough of that, let's look at what's new

I'm trying an experiment this year and using mybinder.org to create an interactive version of the What's New notebook:

https://mybinder.org/v2/gh/rdkit/UGM_2020/master

or <https://tinyurl.com/RDKit-Notebook-2020>

(the link is in the discord chat)

The notebook itself is also in the in the UGM github repo:

https://github.com/rdkit/UGM_2020

You can install the RDKit nightly build for Python 3.7/Linux that I used with:

```
conda install -c rdkit/label/nightly rdkit
```

