MDAnalysis UGM 2023
State of the Union

*Join the discussion on **Discord**:*

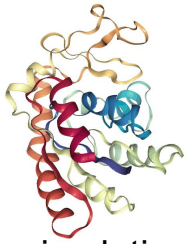
https://discord.com/invite/sAKgZZnPv4

MDANALYSIS

***UGM repository** for relevant material:*

https://github.com/MDAnalysis/UGM2023

An overview of the MDAnalysis library



simulation trajectory

dcd, xtc, trr, ncdf, trj, pdb, pqr, gro, crd, dms, trz, mol2, xyz, config, history, gms, ...

psf, tpr, prmtop, dms, mol2, hoomd xml, ...

MD ANALYSIS

NumPy

"accessible" structured data

`numpy.ndarray()`

python™

analysis algorithm

MD ANALYSIS

tables

images

graphs

processed data

Insights & publication!

JMB

ScienceDirect

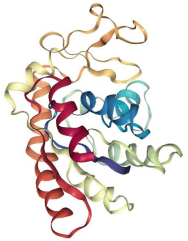
Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open ↔ Closed Transitions

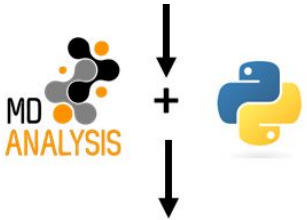
Oliver Beckstein[1,2*†], Elizabeth J. Denning[3†], Juan R. Perilla[3] and Thomas B. Woolf[2,3*]

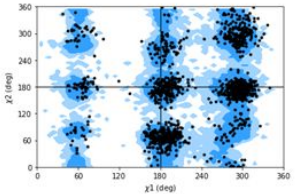
Introduction

The **MDAnalysis** library

- Open source (GPLv2+) **Python** library for handling simulation data
  - Focus on analysing molecular dynamics data
  - … but really any *N*=const particle-based “trajectories”
- Components to build custom analyses and workflows
  - High level: complete analysis classes (RMSD, RMSF, density, dihedrals/Ramachandran, ENCORE, HOLE, *g*(*r*), …)
  - Low level: trajectory data, distance calculations (with PBC), …
- Platform agnostic
  - All major OS (Linux, macOS, Windows)
  - All major CPU architectures
  - All major MD engine file formats







The **MDAnalysis** library

- Support for over 40 file formats
  - **Topologies** (read-only) & **coordinates** (single frame & trajectories)
  - Extensible via *Chemfiles* converter
  - Extensible via own classes (no source code modification necessary)
- MD package independence
  - own internal unit convention (Å, ps ,...)
  - consistent numbering
  - seamless conversion

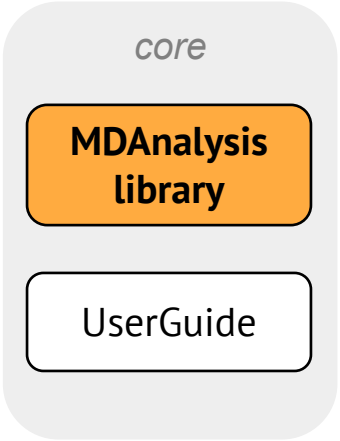
```
import MDAnalysis as mda
u = mda.Universe("in.prmtop", "in.nc")

u.atoms.write("out.xtc", frames="all")
```

Software	File Type
AMBER	PRMTOP, RST7, TRJ, NETCDF
GROMACS	ITP, TPR, GRO, TRR, XTC
CHARMM	PSF, DCD, CRD
NAMD	DCD, COOR, NAMDBIN
LAMMPS	CONFIG, DATA, DUMP, DCD
DL_POLY	CONFIG, HISTORY
HOOMD	XML, GSD
GAMESS	GMS
DESRES	DMS
Others	XYZ, TXYZ, PDB, PDBQT, PQR, TRZ, MOL2, MMTF, FHIAIMS, H5MD, etc...

An overview of the MDAnalysis Organization

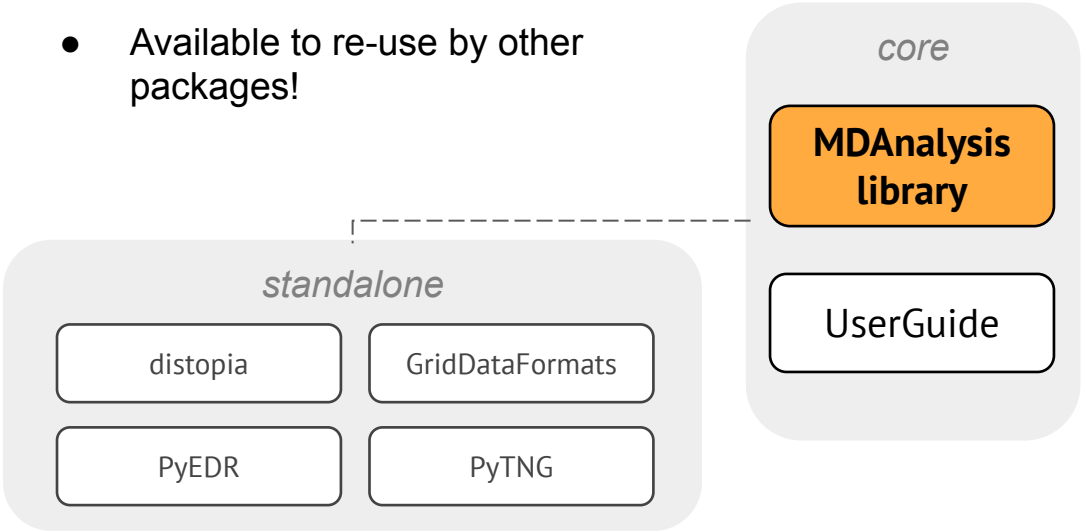
An overview of the MDAnalysis Organization

- **User Guide:** Provides worked examples and information about how to use the MDAnalysis library
- Starting to build out comprehensive developer documentation



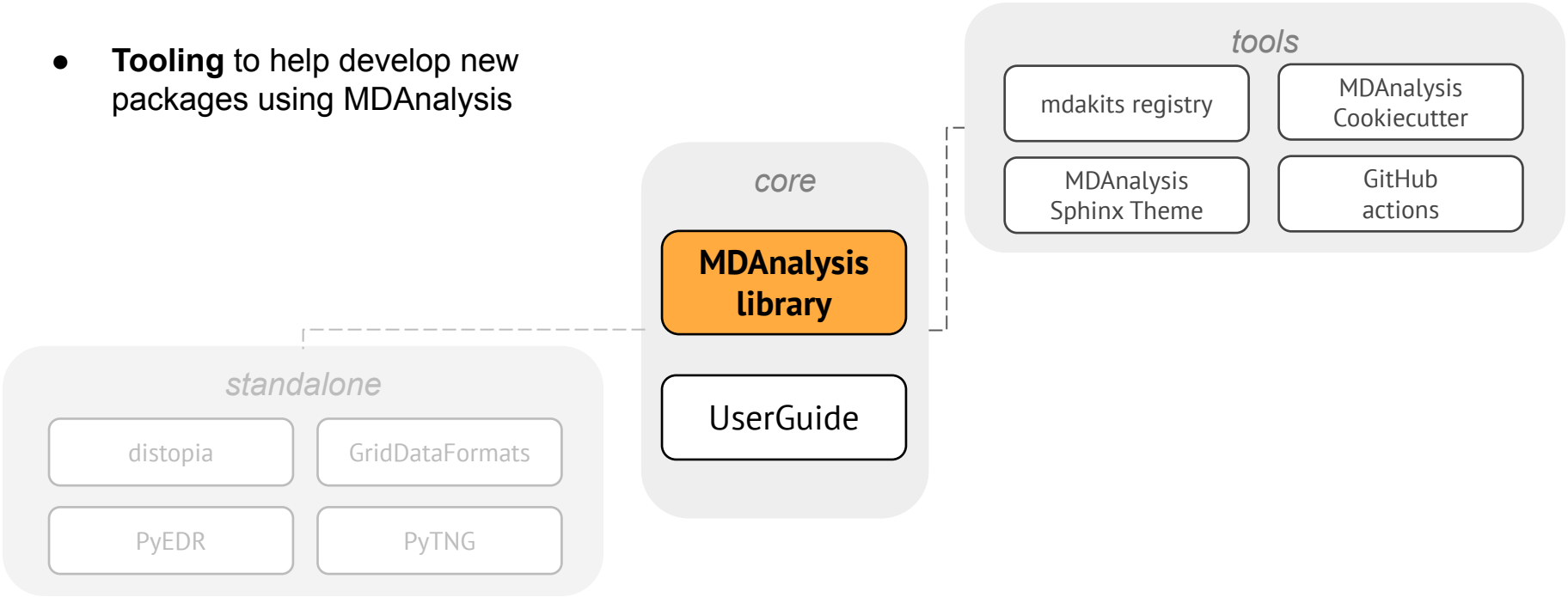
An overview of the MDAnalysis Organization

- **Standalone packages** used by the MDAnalysis library
- Available to re-use by other packages!



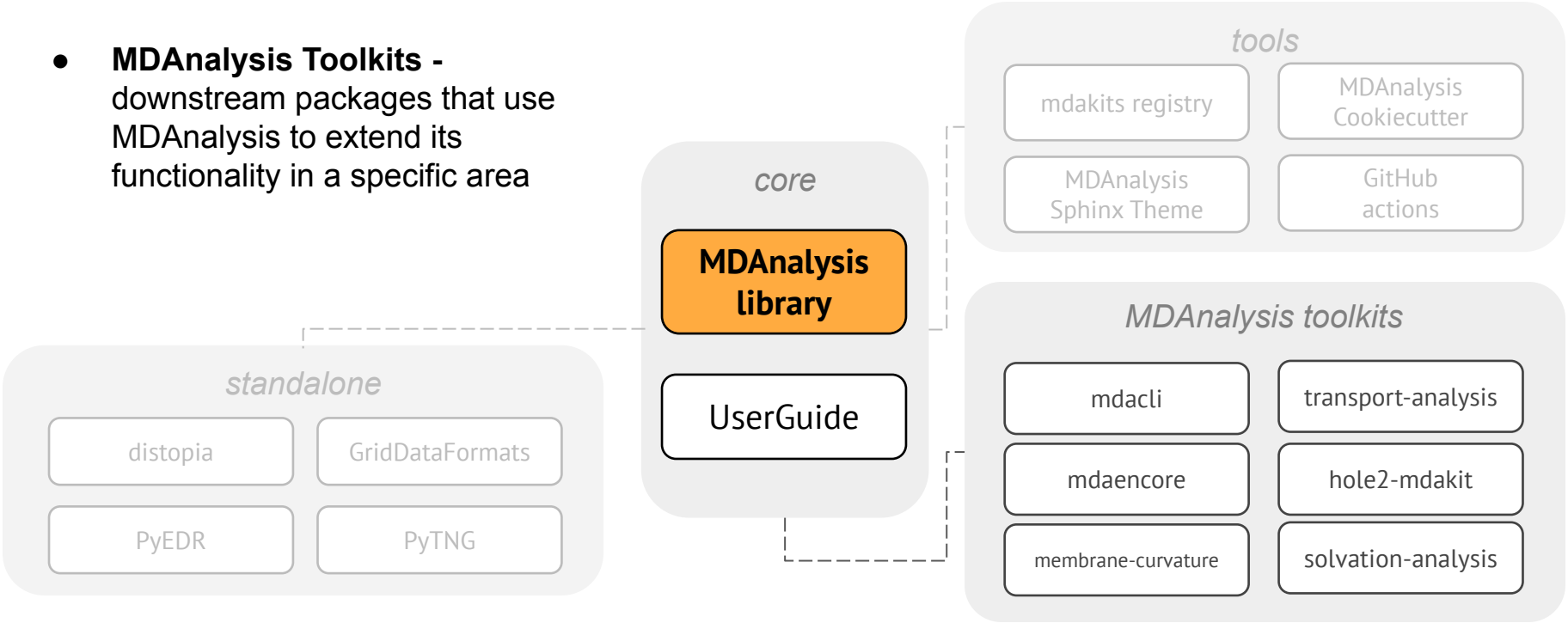
An overview of the MDAnalysis Organization

- **Tooling** to help develop new packages using MDAnalysis



An overview of the MDAnalysis Organization

- **MDAnalysis Toolkits -** downstream packages that use MDAnalysis to extend its functionality in a specific area



An overview of the MDAnalysis Organization

outreach

- mentorship
- workshops

core

- **MDAnalysis library**
- UserGuide

tools

- mdakits registry
- MDAnalysis Cookiecutter
- MDAnalysis Sphinx Theme
- GitHub actions

MDAnalysis toolkits

- mdacli
- transport-analysis
- mdaencore
- hole2-mdakit
- membrane-curvature
- solvation-analysis

standalone

- distopia
- GridDataFormats
- PyEDR
- PyTNG

An overview of the MDAnalysis Organization

outreach

- mentorship
- workshops

core

- **MDAnalysis library**
- UserGuide

tools

- mdakits registry
- MDAnalysis Cookiecutter
- MDAnalysis Sphinx Theme
- GitHub actions

standalone

- distopia
- GridDataFormats
- PyEDR
- PyTNG

MDAnalysis toolkits

- mdacli
- transport-analysis
- mdaencore
- hole2-mdakit
- membrane-curvature
- solvation-analysis

And much more!

Who is MDAnalysis?

188 code contributors and countless community members

Naveen Michaud-Agrawal, Elizabeth J. Denning, Danny Parton, Philip Fowler, **Tyler Reddy**, Joseph Goose, **Jan Domanski**, Benjamin Hall, Paul Rigor, David Caplan, Christian Beckstein (logo), **Sébastien Buchoux**, Joshua L. Adelman, Lukas Grossar, Andy Somogyi, Lukas Stelzl, Jinju Lu, Joshua L. Phillips, Zhuyi Xue, Xavier Deupi, **Manuel Nuno Melo**, Robert McGibbon, Alejandro Bernardin, Lennard van der Feltz, Matthieu Chavent, Joe Jordan, Alex Nesterenko, Caio S. Souza, Sean L. Seyler, **David L. Dotson**, Carlos Yanez S., Kyle J. Huston, Isaac Virshup, **Max Linke**, Gorman Stock, Hai Nguyen, Balasubramanian, Mattia F. Palermo, Utkarsh Saxena, Abhinav Gupta, John Detlefs, Eugen Hruska, Bart Bruininks, Robert Delgado, Wouter Boomsma, **Matteo Tiberti**, Tone Bengtsen, Shantanu Srivastava, Pedro Reis, Ruggero Cortini, Zhiyi Wu, Kashish Punjani, Utkarsh Bansal, Shobhit Agarwal, Vedant Rathore, Akshay Gupta, Juan Eiros Zamora, Jon Kapla, Sang Young Noh, Andrew William King, Kathleen Clark, Dominik 'Rathann' Mierzejewski, Nestor Wendt, **Micaela Matta**, Jose Borreguero, Sören von Bülow, Nabarun Pal, Mateusz Bieniek, Paul Smith, Navya Khare, **Johannes Zeman**, Ayush Suhane, Davide Cruz, Shujie Fan, Andrew R. McCluskey, Henry Mull, **Philip Loche**, Matthew W. Thompson, Ali Ehlen, Daniele Padula, Ninad Bhat, Fenil Suchak, Yibo Zhang, Luís Pedro Borges Araújo, Abhishek A. Kognole, **Rocco Meli**, Matthijs Tadema, Joao Miguel Correia Teixeira, Charlie Cook, Yuanyu Chang, Guillaume Fraux, Ivan Hristov, Michael Quevillon, Hao Tian, **Hugo MacDermott-Opeskin**, Anshul Angaria, Shubham Sharma, Yuxuan Zhuang, Cédric Bouysset, Abhishek Shandilya, Morgan L. Nance, Faraaz Shah, Wiep van der Toorn, Siddharth Jain, Ameya Harmalkar, Shakul Pathak, Andrea Rizzi, William Glass, Marcello Sega, Edis Jakupovic, Nicholas Craven, Mieczyslaw Torchala, Ramon Crehuet, Haochuan Chen, Karthikeyan Singaravelan, Ian M. Kenney, Aditya Kamath, Leonardo Barneschi, Henrik Jäger, Jan Stevens, Orion Cohen, Dimitrios Papageorgiou, Hannah Pollak, Estefania Barreto-Ojeda, Paarth Thadani, Henry Kobin, Kosuke Kudo, Sulay Shah, Alexander Yang, Filip T. Szczypiński, Marcelo C. R. Melo, Mark D. Driver, Kevin Boyd, Atharva Kulkarni, Yantong Cai, Bjarne Feddersen, Pratik Gupta, Alexander Gorfer, Aya M. Alaa, Kazi Shudipto Amin, Alia Lescoulie, Henok Ademtew, Uma D Kadam, Tamandeep Singh, Mingyi Xue, Meghan Osato, Anirvinya G, Rishabh Shukla, Manish Kumar, Aditi Tripathi, Sukeerti T, Kavya Bisht, Mark Verma, Marcelo D. Poleto, Ricky Sexton, Rafael R. Pappalardo, Tengyu Xie, Raymond Zhao, Haleema Khan, Jennifer A Clark, Jake Fennick, Utsav Khatu, Patricio Barletta, Mikhail Glagolev, Christian Pfaendner, Pratham Chauhan, Meet Brijwani, Vishal Parmar, Moritz Schaeffler, Xu Hong Chen, Domenico Marson, Ahmed Salah Ghoneim, Alexander Schlaich, Josh Vermaas, Xiaoxu Ruan, Egor Marin, Shaivi Malik, Daniel J. Evans, Mohit Kumar, Shubham Kumar, Zaheer Timol, Geongi Moon

MDAnalysis Personnel

(here at the UGM!)

Core Developers

(unable to attend)

Fiona
@fiona-naughton

Irfan
@IAlibay

Lily
@lilyminium

Hugo
@hmacdope

Rocco
@RMeli

Micaela
@micaela-matta

Richard
@richardjgowers

Oliver
@orbeckst

Tyler
@tylerjreddy

Project /Community manager

Jenna
@jennaswa

Extended MDA Team @UGM

Ian
@ianmkenney

Yuxuan
@yuxuanzhuang

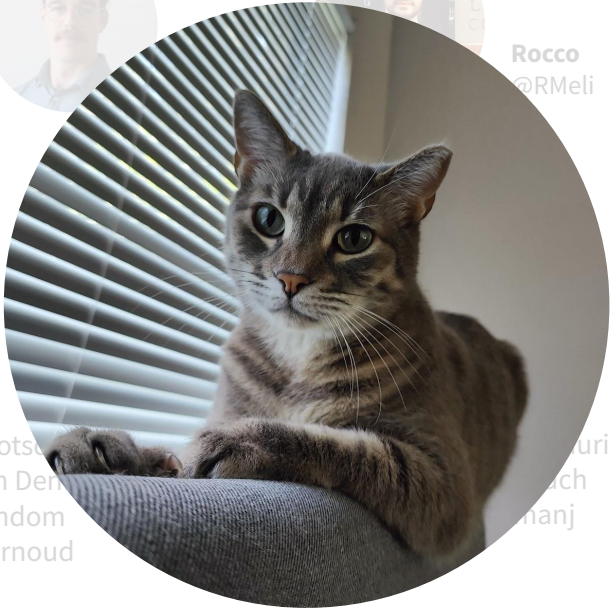
Emeriti Core Devs

@dotsdl
Elizabeth Denning
@jandom
@jbarnoud

@kain88-de
@mnmelo
@mtiberti
@nmichaud

@PicoCentauri
@seb-buch
@zemanj

MDAnalysis Personnel: Emotional Support

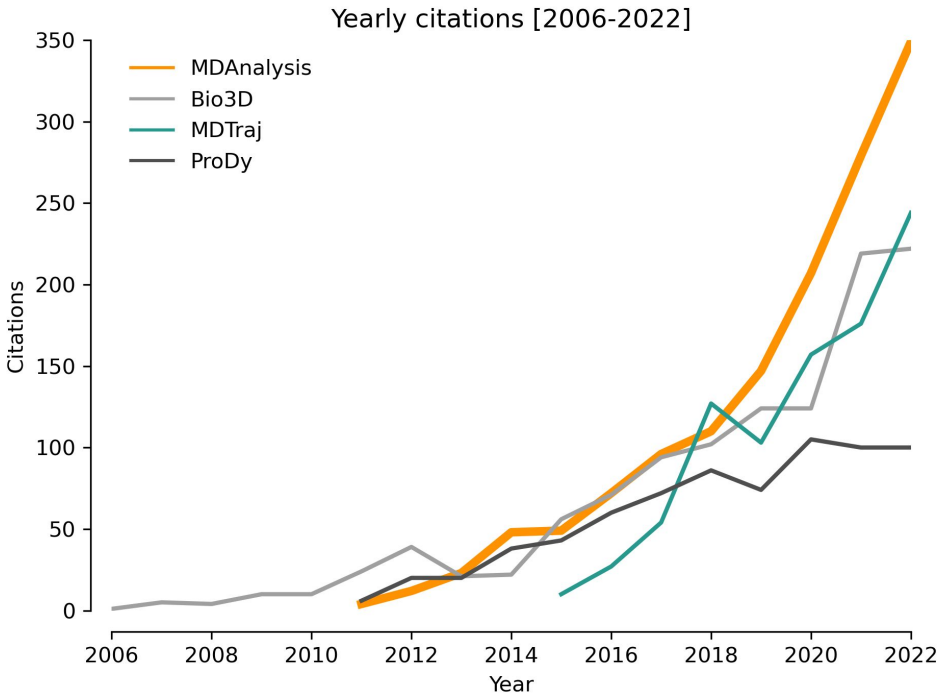


Health of the Project

How are we doing?

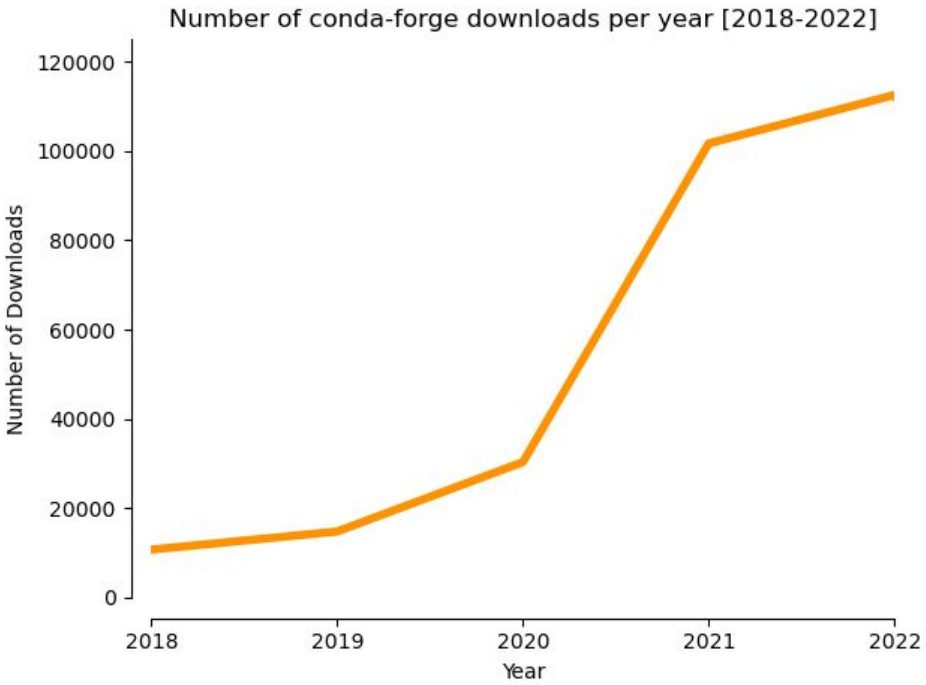
(focusing on the core library)

Health of the Project: Citations



- Scopus citation search of MDAnalysis references
- Caveat: does not index all sources (e.g. JOSS)

Health of the Project: Downloads



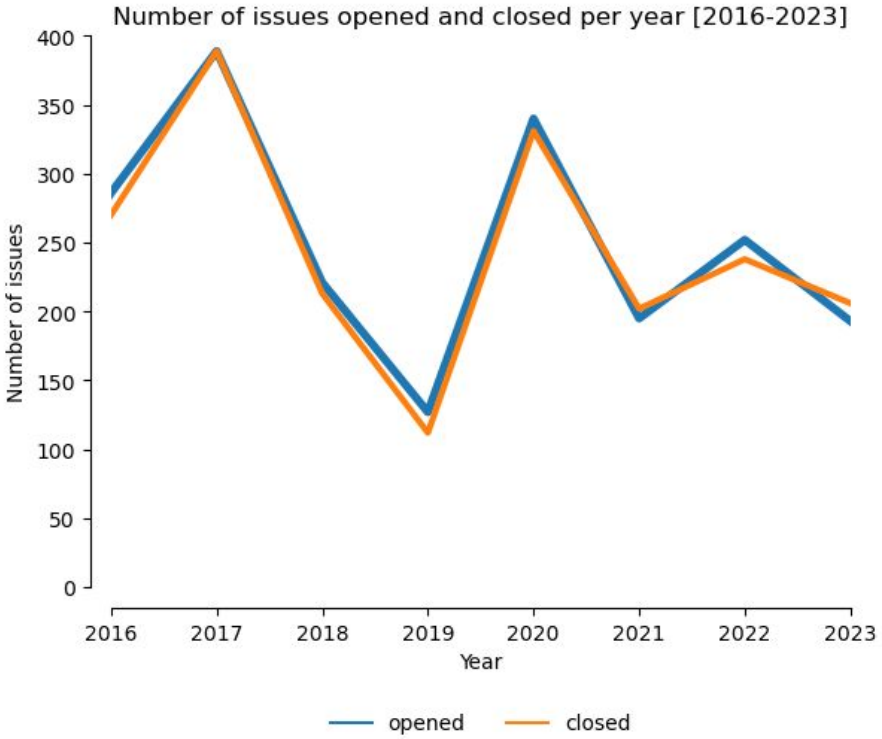
Conda downloads:

- Yearly downloads over 2018 to 2022
  - Obtained through condastats
- Caveats
  - Includes CI, bots, etc..
  - Concurrent with increasing reliance on conda use

Other metrics:

- 22430 PyPI monthly downloads
- 120 downstream packages (GitHub)

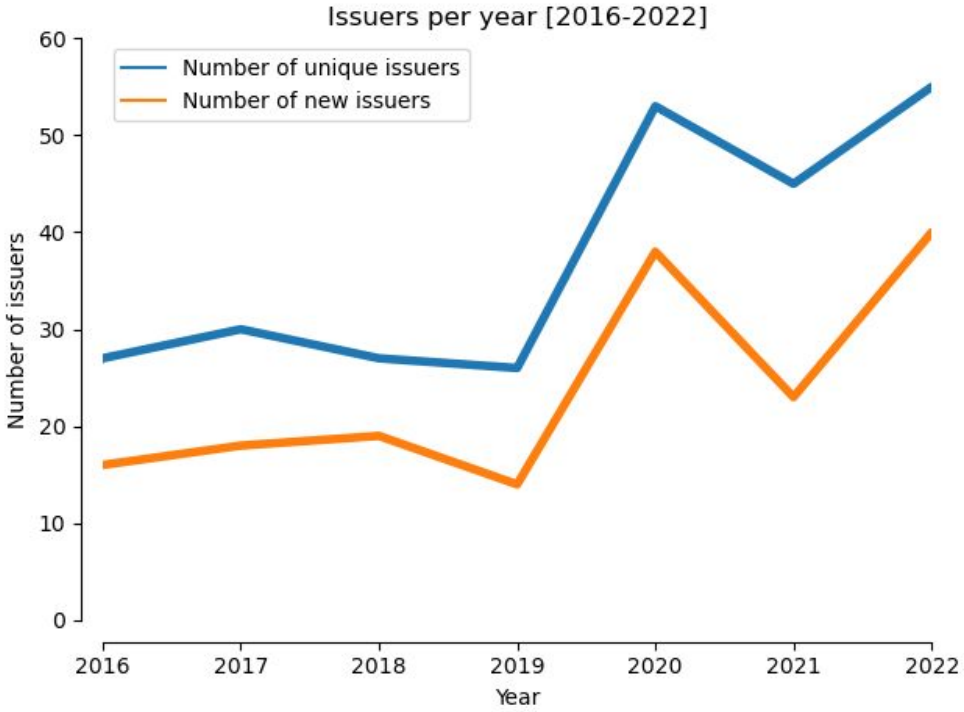
Health of the Project: Issues



410 issues currently open

- Average issue retention time of **37 days**
- Long-standing issues causes:
  - Lower priority features
    - “Nice to have”
  - Obscure bugs
    - Hard to debug
    - Lack of expertise
  - API break needing changes
  - Known limitations

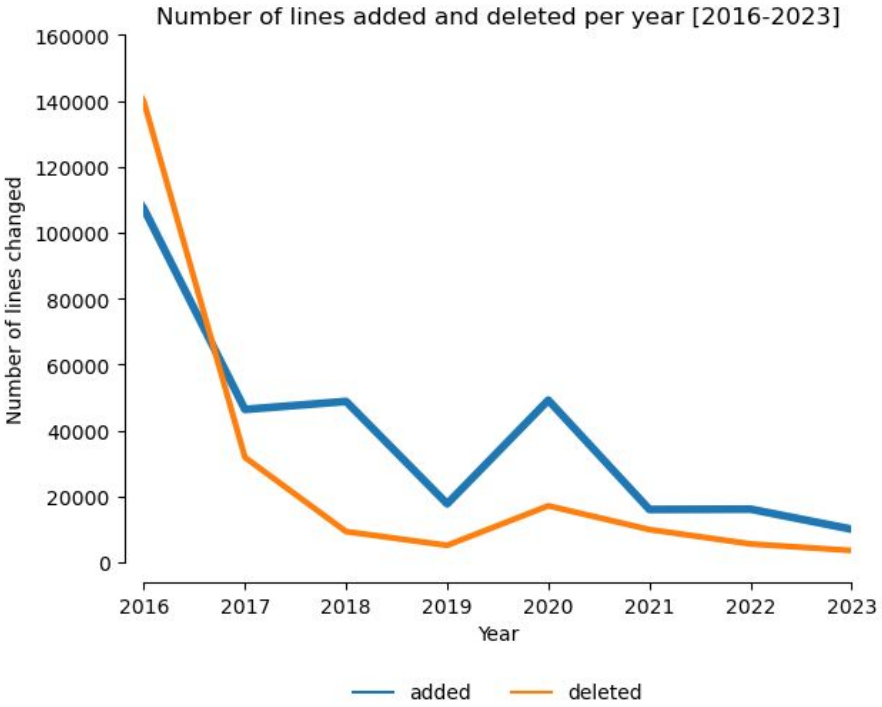
Health of the Project: Issues



Using issues as a metric for activity

- Unique issuers:
  - Per year, the number of unique individuals raising issues
- New issuers
  - Issuers who have never made an issue before
- Slowly increasing amount of activity
  - Small proportion of total issues raised

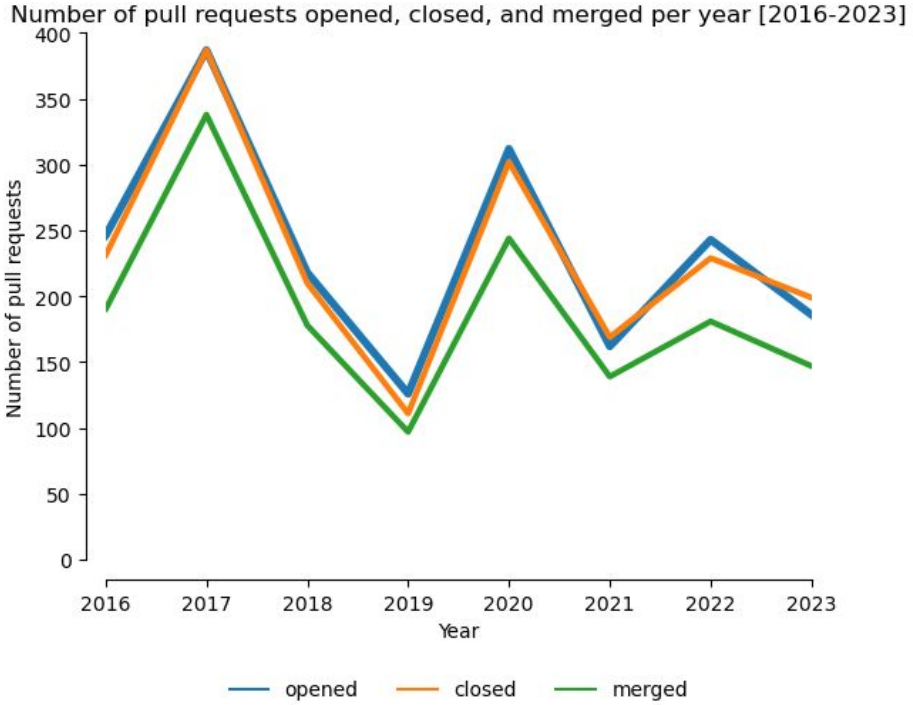
Health of the Project: Contributions



Activity based on code additions / deletions

- Note: change in which files get merged circa 2016
- Apparent slowdown in volume of code contributions
- Potential causes
  - Shift towards contributing to other packages in ecosystem (standalone tools & MDAKits)
  - Increasing maintenance overhead
  - Changes in API stability priorities post v1.0 release (~ 2020)

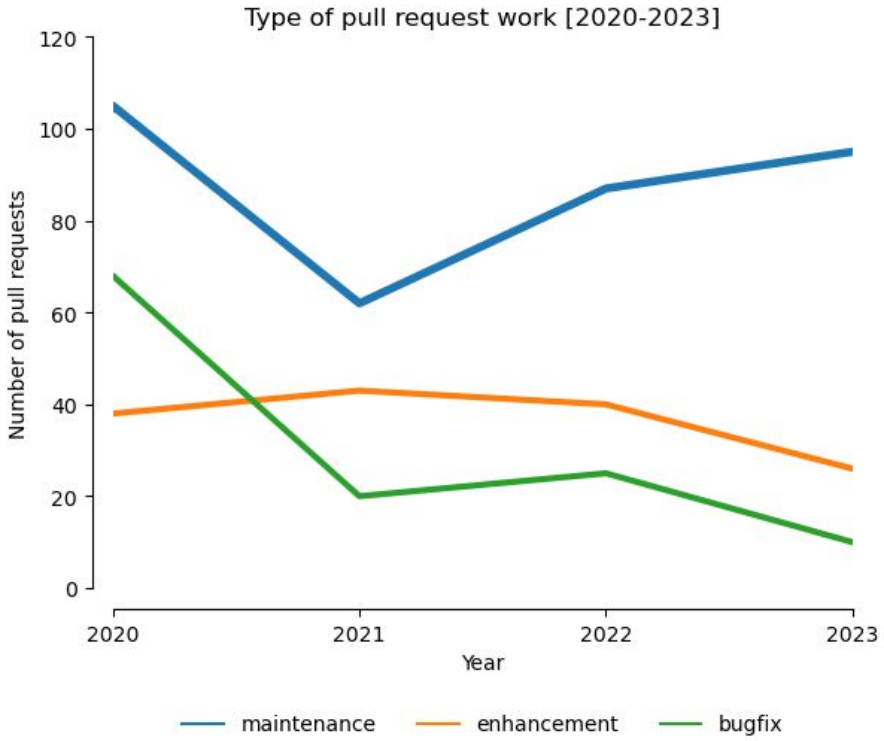
Health of the Project: Contributions - PRs



Pull Requests are a direct indicator of developer activity

- Most PRs make it to merged status
- Mostly sustained but slightly declining rate of contribution
- Peaks of activity near major releases (e.g. v1.0 in 2020)

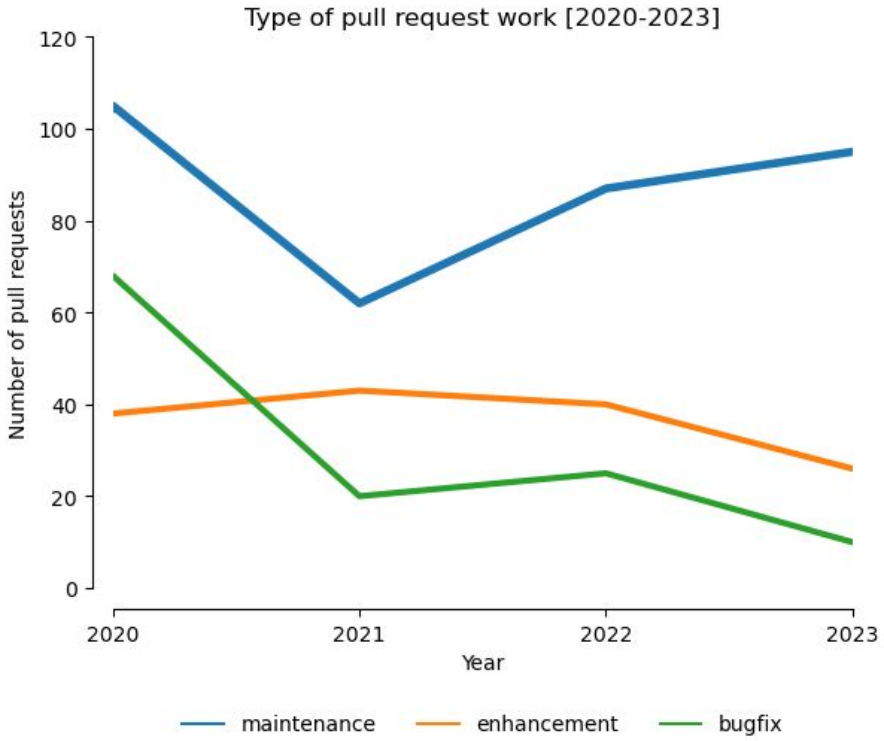
Health of the Project: Contributions - PRs



Breaking down contribution by type

- Maintenance
  - Work to keep up with ecosystem changes, continuous integration, deployment, etc..
- Enhancement
  - Addition of new features or improvement of existing ones
- Bugfix
  - Fixing pre-existing issues in codebase

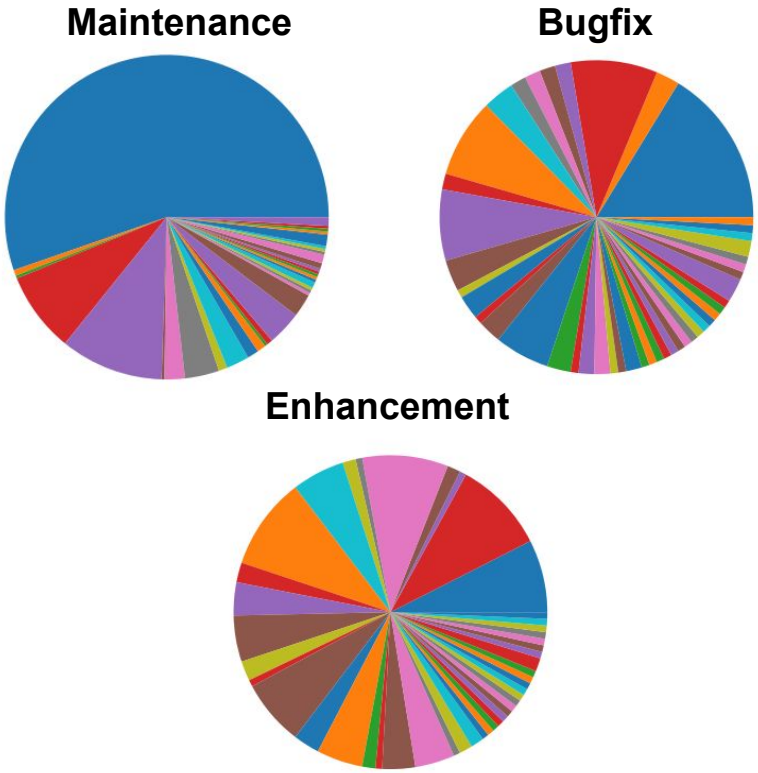
Health of the Project: Contributions - PRs



Breaking down contribution by type

- Note peak in 2020 due to v1.0 and 2.0 releases
  - Lots of historical code removals counted as maintenance
- Seeing a steady increase in maintenance over time
  - Ever rapidly changing ecosystem
  - Support for more OS, hardware, etc..
  - More code leads to more maintenance

Health of the Project: Contributions - PRs



Developer contribution diversity by type (2020-2023)

- Generally quite diverse contributor set for bug fixes and enhancements
- Maintenance tends to fall to a smaller set of contributors
  - ~ 75% of all contributions by 3 developers
  - Tend to be less glamorous & advertised tasks
    - Releases, CI, packaging, etc…

Health of the Project: Funding

Chan Zuckerberg Initiative grants:

*EOSS4: Faster, Extensible Molecular Analysis for Reproducible Science (**2022**)*
*EOSS5: Growing the MDAnalysis community sustainably (**2023**)*

NSF CSSI Elements (upcoming):

*Streaming Molecular Dynamics Simulation Trajectories for Direct Analysis (**2023+**)*

Health of the Project: Funding

Smaller grants & funding sources:

NumFOCUS small development grants

- Up to $10,000, call opens three times a year
- Looking for project ideas!

NUMFOCUS
OPEN CODE = BETTER SCIENCE

Google Summer of Code / Season of Docs

- Supports 1-2 new contributor projects per year

Some key ongoing work

MDAnalysis has been quite busy over the years!

Some key ongoing areas of work:

- Towards a **faster** library
  - Low level code optimization
  - Analysis parallelisation

- Towards a **slimmer, more maintainable**, library
  - Building an ecosystem of packages through MDAKits
  - Migrating difficult to maintain codes out of the core library
  - License changes

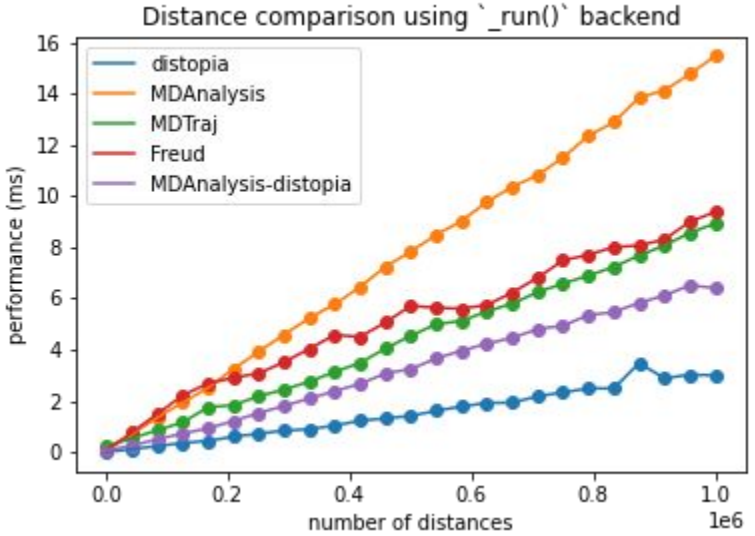
Towards a **faster** MDAnalysis: Cythonization

- Overcoming Python limitations
  - **Poor memory access** to underlying NumPy arrays
  - Hard to optimally leverage hardware features

- Ongoing work to **Cythonize key data structures**
  - Better C/C++ interface (libmdanalysis)
  - Timestep and other coordinate handling objects
  - Towards fast Cythonized readers & writers

Towards a **faster** MDAnalysis: Distopia

- Stand-alone replacement for aging `MDAnalysis.lib.distances`
- Heavily leverages **vector instructions**
- Showing a performance improvement of **4-10x**
- Currently an optional backend to library (v2.5+), will automatically switch in the future

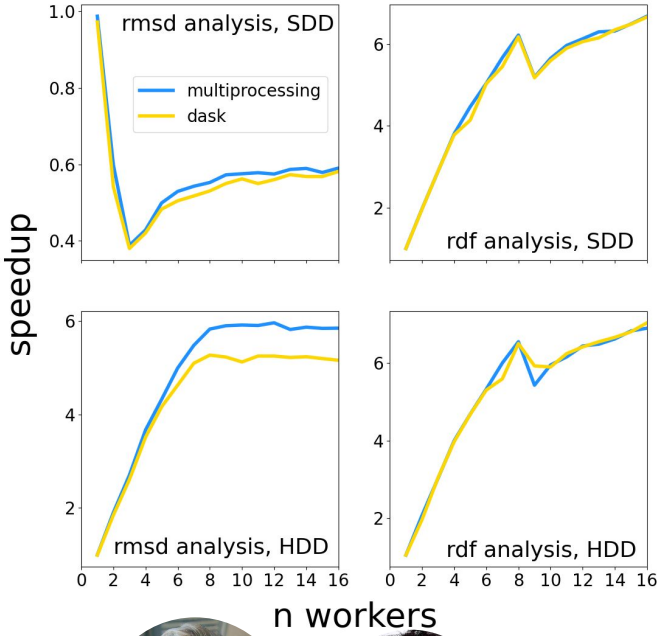


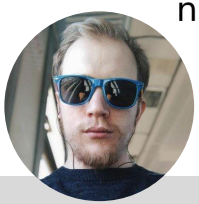
https://github.com/MDAnalysis/distopia

Towards a **faster** MDAnalysis: Parallelisation

Leverage **multi-core parallelism** for analysis methods

- Enable analysis methods to directly leverage multicore parallelism backends (e.g. Dask and Multiprocessing)
  - Including cluster support
  - Quasi-invisible to users
- **5-8x** speedup when not IO-bound (e.g. SSD)



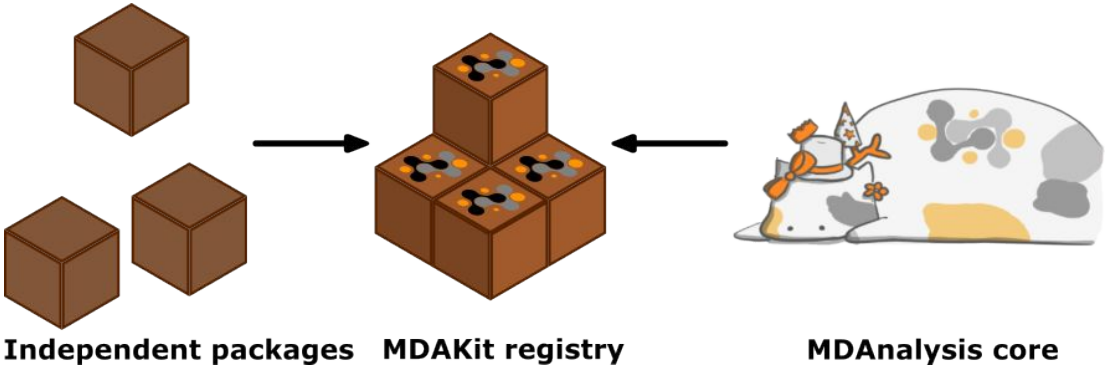


See Egor's talk tomorrow afternoon!

Towards a **slimmer** MDAnalysis library

- Core library maintenance is becoming increasingly burdensome
  - Rapid Python & upstream release schedule (e.g. NEP29)
  - Many hardware flavours to support
- Adding new features is a slow and intensive process
  - Need to keep to strict packaging rules
  - Limited core developer knowledge
- Looking to de-bloat the MDAnalysis library by encouraging an ecosystem of MDAnalysis-using packages

MDAnalysis Toolkits (MDAKits)



MDAKit	Keywords	Authors	CI badges
hole2-mdakit	pores, ion channels, transporters, HOLE	hole2-mdakit authors	latest passed, develop passed
lipyds	lipids, membranes	lilyminium	latest passed, develop unavailable
mda-tui	transformations, TUI	mda-tui authors	latest passed, develop passed
mdachecker	structure analysis, conformation checks	lAlibay	latest passed, develop passed
mdacli	command line interface, molecular-dynamics	mdacli authors	latest passed, develop passed
mdaencore	ensemble, similarity, covariance, PCA	mdaencore authors	latest passed, develop passed
openmm-mdanalysis-reporter	OpenMM, Reporters	openmm-mdanalysis-reporter authors	latest passed, develop passed

Providing tooling and documentation to support package developers

- Cookiecutter and Registry

See Ian's talk tomorrow morning!

https://mdakits.mdanalysis.org/mdakits.html

MDAnalysis' new MDAKits

- ***mdacli:*** a command-line interface to MDAnalysis Analysis classes
- ***solvation-analysis:*** methods for analyzing the solvation structure of liquids
- ***membrane-curvature:*** analysis of membrane curvatures
- ***transport-analysis:*** tools for computing and analyzing transport properties

Philip & Joao
@picocentauri
@joaomcteixeira

Orion
@orionarcher

Estefania
@ojeda-e

Xu Hong
@xhgchen

Upcoming library changes

- Lowering the number of core dependencies
  - allowing smaller minimal packages
  - networkX, matplotlib, biopython

- Moving harder to maintain components to downstream packages
  - encore
  - HOLE2
  - waterdynamics
  - Converters
  - .. others?

mdaencore

hole2-mdakit

Easier downstream development: relicensing!

- Current GNU Public License v2+ too restrictive
  - Copyleft applied on `import MDAnalysis`
  - Restricts licenses you can release under

- Starting the slow process of relicensing to LGPL v2.1+
  - Allow freedom of import
  - Retain copyleft for direct code changes
  - Will be contacting all historical developers
  - Email licensing@mdanalysis.org for questions

Read our latest blog post!

Other ongoing works

- Interoperability through converters
  - OpenMM, **RDKit**, Parmed
  - *Coming soon:* ASE, MDTraj, and OpenFF Tk
- Better guessers
  - Inferring from file format information (e.g. PDB)
  - Provide better clarity on guessed attributes
- New readers
  - New **auxiliary** readers (EDR, etc..)
  - Large XTC formats, various LAMMPS improvements, H5MD…

Cedric
@cbouy

See Cedric's talk tomorrow!

Aya
@aya9aladdin

Bjarne
@BFedder

Future plans: v3.0 and beyond 🚀

Join us at **tomorrow at 4 pm** for a discussion on MDAnalysis' future!

Some potential ideas:

- New data processing paradigms
  - cloud streaming, faster bond handling, …
- Support for new analyses
  - DSSP, SASA, etc..
- Better chemical perception & consistency
  - Unit handling via *pint*, improved guessers, better cheminformatics
- Your own needs and ideas!

Community building and engagement

MDAnalysis is not just about code!

- Mentorship programs
- Upcoming workshops!
  - October 25th - *sold out*
  - More later in the year (next ~ mid November)
  - Looking for tutors and partner projects!
- Next UGM in **2024**!

CompChemURG
(The Binding Site)

Getting involved

MDAnalysis is always looking for new contributors!

- Join in on user discussions and meetings!
- Submit issues
- Tackle of our many issues
- Create your own MDAKit
- Teach a workshop
- Participate in a mentorship program

See Friday's Hackathon session for various ways to get started with contributing!

https://github.com/MDAnalysis/UGM2023/tree/main/hackathon

Acknowledgements

188 code contributors and countless community members

Naveen Michaud-Agrawal, Elizabeth J. Denning, Danny Parton, Philip Fowler, **Tyler Reddy**, Joseph Goose, **Jan Domanski**, Benjamin Hall, Paul Rigor, David Caplan, Christian Beckstein (logo), **Sébastien Buchoux**, Joshua L. Adelman, Lukas Grossar, Andy Somogyi, Lukas Stelzl, Jinju Lu, Joshua L. Phillips, Zhuyi Xue, Xavier Deupi, **Manuel Nuno Melo**, Robert McGibbon, Alejandro Bernardin, Lennard van der Feltz, Matthieu Chavent, Joe Jordan, Alex Nesterenko, Caio S. Souza, Sean L. Seyler, **David L. Dotson**, Carlos Yanez S., Kyle J. Huston, Isaac Virshup, **Max Linke**, Gorman Stock, Hai Nguyen, Balasubramanian, Mattia F. Palermo, Utkarsh Saxena, Abhinav Gupta, John Detlefs, Eugen Hruska, Bart Bruininks, Robert Delgado, Wouter Boomsma, **Matteo Tiberti**, Tone Bengtsen, Shantanu Srivastava, Pedro Reis, Ruggero Cortini, Zhiyi Wu, Kashish Punjani, Utkarsh Bansal, Shobhit Agarwal, Vedant Rathore, Akshay Gupta, Juan Eiros Zamora, Jon Kapla, Sang Young Noh, Andrew William King, Kathleen Clark, Dominik 'Rathann' Mierzejewski, Nestor Wendt, **Micaela Matta**, Jose Borreguero, Sören von Bülow, Nabarun Pal, Mateusz Bieniek, Paul Smith, Navya Khare, **Johannes Zeman**, Ayush Suhane, Davide Cruz, Shujie Fan, Andrew R. McCluskey, Henry Mull, **Philip Loche**, Matthew W. Thompson, Ali Ehlen, Daniele Padula, Ninad Bhat, Fenil Suchak, Yibo Zhang, Luís Pedro Borges Araújo, Abhishek A. Kognole, **Rocco Meli**, Matthijs Tadema, Joao Miguel Correia Teixeira, Charlie Cook, Yuanyu Chang, Guillaume Fraux, Ivan Hristov, Michael Quevillon, Hao Tian, **Hugo MacDermott-Opeskin**, Anshul Angaria, Shubham Sharma, Yuxuan Zhuang, Cédric Bouysset, Abhishek Shandilya, Morgan L. Nance, Faraaz Shah, Wiep van der Toorn, Siddharth Jain, Ameya Harmalkar, Shakul Pathak, Andrea Rizzi, William Glass, Marcello Sega, Edis Jakupovic, Nicholas Craven, Mieczyslaw Torchala, Ramon Crehuet, Haochuan Chen, Karthikeyan Singaravelan, Ian M. Kenney, Aditya Kamath, Leonardo Barneschi, Henrik Jäger, Jan Stevens, Orion Cohen, Dimitrios Papageorgiou, Hannah Pollak, Estefania Barreto-Ojeda, Paarth Thadani, Henry Kobin, Kosuke Kudo, Sulay Shah, Alexander Yang, Filip T. Szczypiński, Marcelo C. R. Melo, Mark D. Driver, Kevin Boyd, Atharva Kulkarni, Yantong Cai, Bjarne Feddersen, Pratik Gupta, Alexander Gorfer, Aya M. Alaa, Kazi Shudipto Amin, Alia Lescoulie, Henok Ademtew, Uma D Kadam, Tamandeep Singh, Mingyi Xue, Meghan Osato, Anirvinya G, Rishabh Shukla, Manish Kumar, Aditi Tripathi, Sukeerti T, Kavya Bisht, Mark Verma, Marcelo D. Poleto, Ricky Sexton, Rafael R. Pappalardo, Tengyu Xie, Raymond Zhao, Haleema Khan, Jennifer A Clark, Jake Fennick, Utsav Khatu, Patricio Barletta, Mikhail Glagolev, Christian Pfaendner, Pratham Chauhan, Meet Brijwani, Vishal Parmar, Moritz Schaeffler, Xu Hong Chen, Domenico Marson, Ahmed Salah Ghoneim, Alexander Schlaich, Josh Vermaas, Xiaoxu Ruan, Egor Marin, Shaivi Malik, Daniel J. Evans, Mohit Kumar, Shubham Kumar, Zaheer Timol, Geongi Moon