SIMCODES Orientation

Ryan M. Richard

Ames National Laboratory and Iowa State University

Disclaimer

- SIMCODES takes the topics in this presentation seriously.
- I am also aware that these topics are incredibly dry.
- To lighten the mood of the orientation I have attempted to add a little humor and some pop culture references.
- This “lighter mood” should NOT be taken as downplaying the importance of these topics.
- Failure to follow the procedures and rules in this presentation will have very unfunny results, including up to removal from SIMCODES.

Topics

- Abbreviations: The Rosetta stone to what I say.
- Objectives and deliverables: What we're all here to do.
- Safety: How to not die while working towards "the goal."
- Expectations: What you must do if you want to be paid.
- Code of conduct: How you should behave.
- Research integrity: How you are expected to conduct research.
- Getting help: What to do if (when?) problems arises.

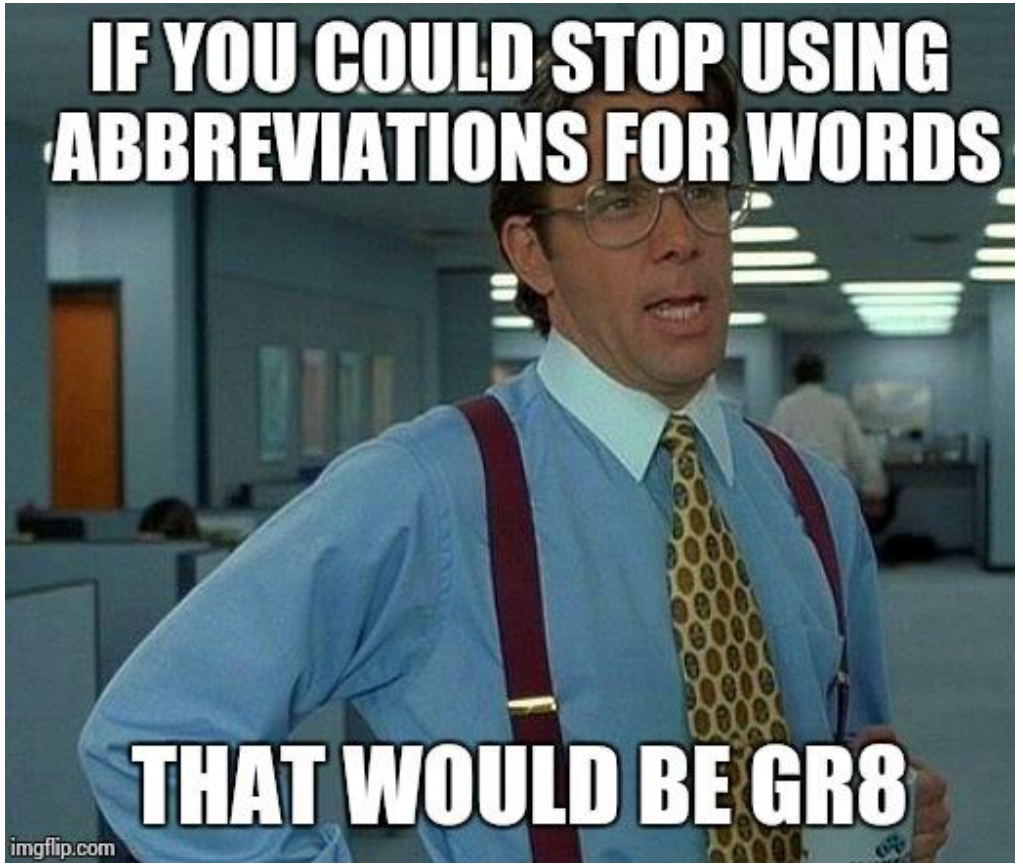
Abbreviations

Abbreviations

- We love to use abbreviations.
- Throughout this program you will be exposed to a lot of abbreviations.
  - Partly because the federal government relies heavily on them, and we are federally funded.
  - Also, partly because computational chemistry relies heavily on them.
- We will do our best to define abbreviations as they come up.
  - If we miss one, ask!
- Let’s get some common ones established…

Abbreviations

- SIMCODES technically stands for “Sustainability Institute for Machine learning and Collaborative Open-source Development of Enzymatic Simulations”, but even I had to look that up…
- ISU is “Iowa State University”.
- DOE is the Department of Energy.
- AMES, or AL, are short for “Ames National Laboratory”
  - Do NOT use ANL (that’s Argonne National Laboratory’s designated abbreviation).
- NSF is short for the “National Science Foundation” (they’re funding SIMCODES).
- REU stands for “Research Experiences for Undergraduates” (that’s the program type of SIMCODES).



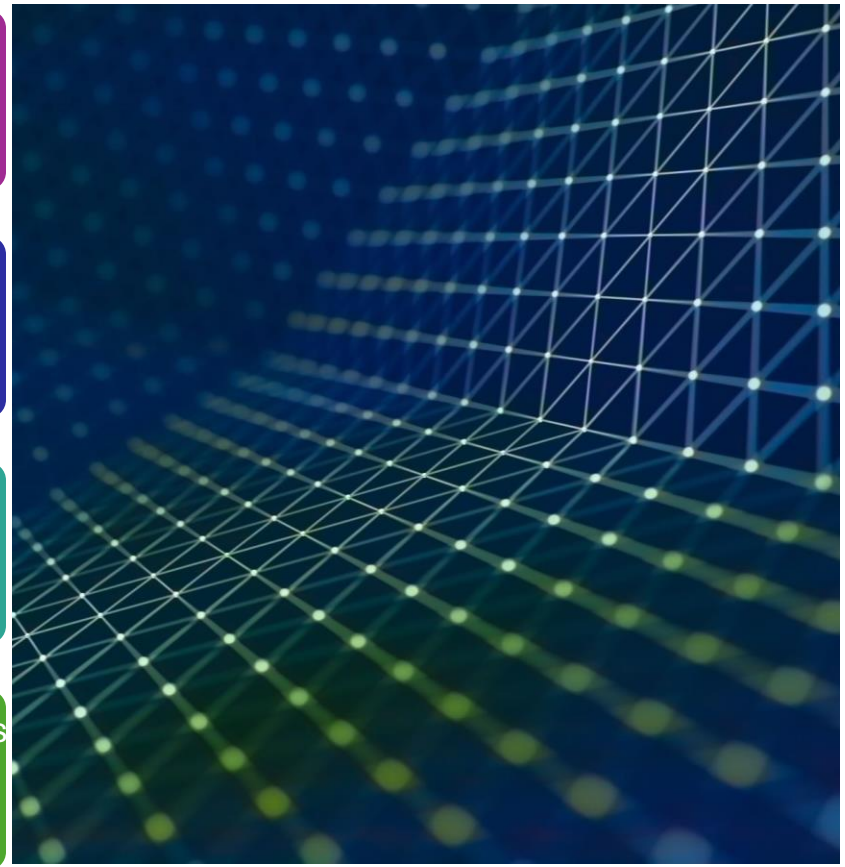
Abbreviations Cont.

- SULI stands for “Science Undergraduate Learning Internship”. It is roughly DOE’s equivalent of the REU program.
  - AMES and ISU have a number of summer activities that target both SULI and REUs, so you’ll likely meet SULI students.
- AI stands for artificial intelligence.
- ML stands for machine learning.
- AI/ML stands for the fact that most people no longer want to distinguish between AI and ML and now treat them as the same thing.

Objectives and Deliverables

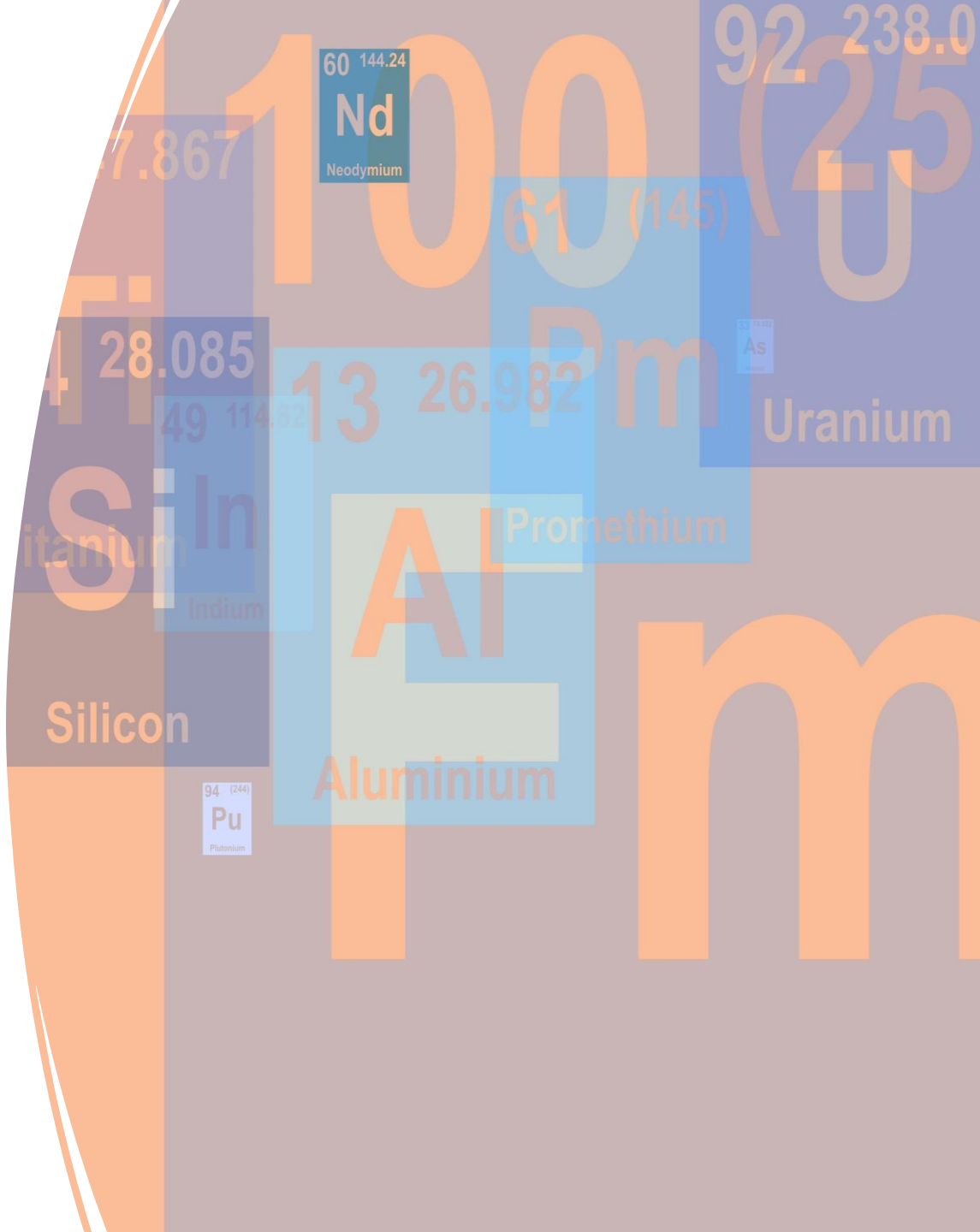
SIMCODES Objectives

Provide	Provide undergraduates with new research experiences in computational chemistry, biochemistry, and computer science.
Provide	Provide hands on, real world, experience with software, simulation, and AI/ML applied to designing next generation enzymatic catalysts.
Provide	Provide research opportunities to undergraduates who may lack such opportunities.
Provide	Provide training on research software best practices to better educate the next generation of computational STEM researchers.



Deliverables

- There will be a midterm presentation June 10.
  - 10-15 minute presentation on your research so far and what you still want to accomplish.
  - Opportunity to practice presenting and to better understand what your colleagues are doing.
- There will be a poster presentation August 1.
  - All students must present a poster.
- There is no obligation to publish results (but we hope you get to!).



Safety

General Safety

- We take safety very seriously!!!
  - You MUST comply with all safety instructions and alarms.
  - If you are unsure how to do something safely, ask!
  - AMES and ISU know there are a lot of visitors on campus for the summer, i.e., expect there to be drills to test that I did my job.
- For emergencies call 911.
- For non-emergencies:
  - In ISU spaces call ISU Police: 515-294-4428 (dpsinfo@iastate.edu)
  - In AMES spaces call guards: 515-294-3483 (guards@ameslab.gov)

(This is a meme. 911 is the only emergency number you will need.)

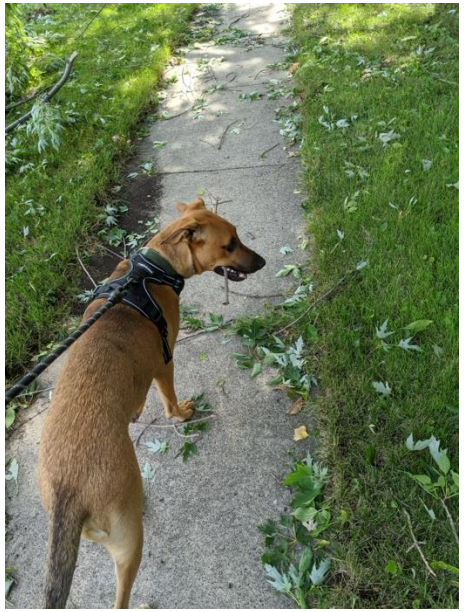
Fire

- If you spot a fire:
  - Attempt to fight the fire only if you are trained (and only if it is still small).
  - If you can not control the fire, and the alarm has not been activated, pull a fire alarm and evacuate.
- If the fire alarm in your building goes off immediately evacuate via the closest non-blocked exit.
  - Leave all belongings behind.
  - Close doors on your way out, if you can, but your safety comes first!

(Again, this is a meme. In a real fire, everything is not fine because there's a freaking fire!)

Severe Weather

- The city of Ames tests the storm sirens the first Wednesday of every month at 10 AM.
  - These tests are not drills and you may continue to work through them.
- In Iowa primary threats to safety are hail, strong winds, and tornadoes.
- If a storm watch/warning is issued you will receive notifications from AMES guards to your AMES email.
- Usually go to the lowest level away from windows.



Research Safety

- Computational internship, so shouldn't be hazards in your workspace, but…
- Most of SIMCODES will take place in chemistry buildings that also house active chemistry labs.
  - Only enter research spaces with permission.
  - Always use the appropriate personal protection equipment (PPE) in such spaces.
  - If you see a spill anywhere, or smell a chemical, immediately report it.
  - Know the location of safety showers and eye wash stations.
- Usual ergonomics best practices apply in computational spaces:
  - Take eye breaks, make sure you move around, sit up straight, etc.
  - If your workspace is uncomfortable bring it up!!!

Cybersecurity

- AMES and ISU email addresses are targets for phishing.
- Be suspicious of any email asking for passwords, account numbers, Social Security number, etc.
  - Be especially suspicious of close deadlines, and threatening emails.
- It is strongly recommended you do not click on links in emails unless you trust the sender.
- When in doubt go to the website yourself (via Google).
- AMES will likely conduct a phishing exercise over the course of the summer. Please don’t fall for it.
  - Report phishing attempts sent to your AMES email to abuse@ameslab.gov.

Additional Notes on Safety

- Safety is a priority for everyone.
- You must know safety procedures for your space.
  - If your mentor does not volunteer them, ask!
  - Again, there will more than likely be drills this summer and you need to know where to go!
- AMES may require you to participate in an accountability exercise.
  - If you get an accountability email, respond ASAP!
  - There is usually a notification about this a week or so in advance.

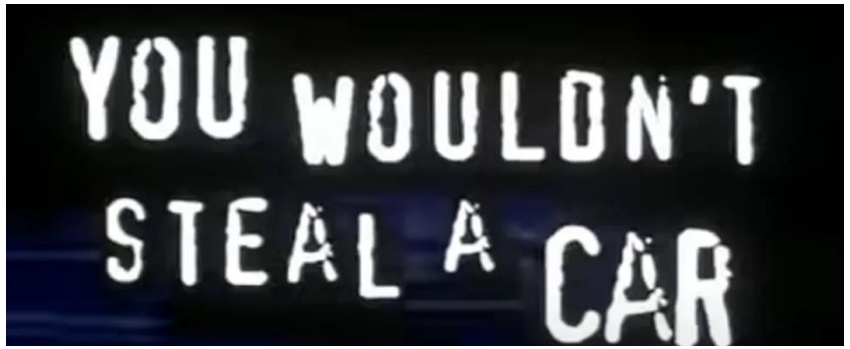
Expectations

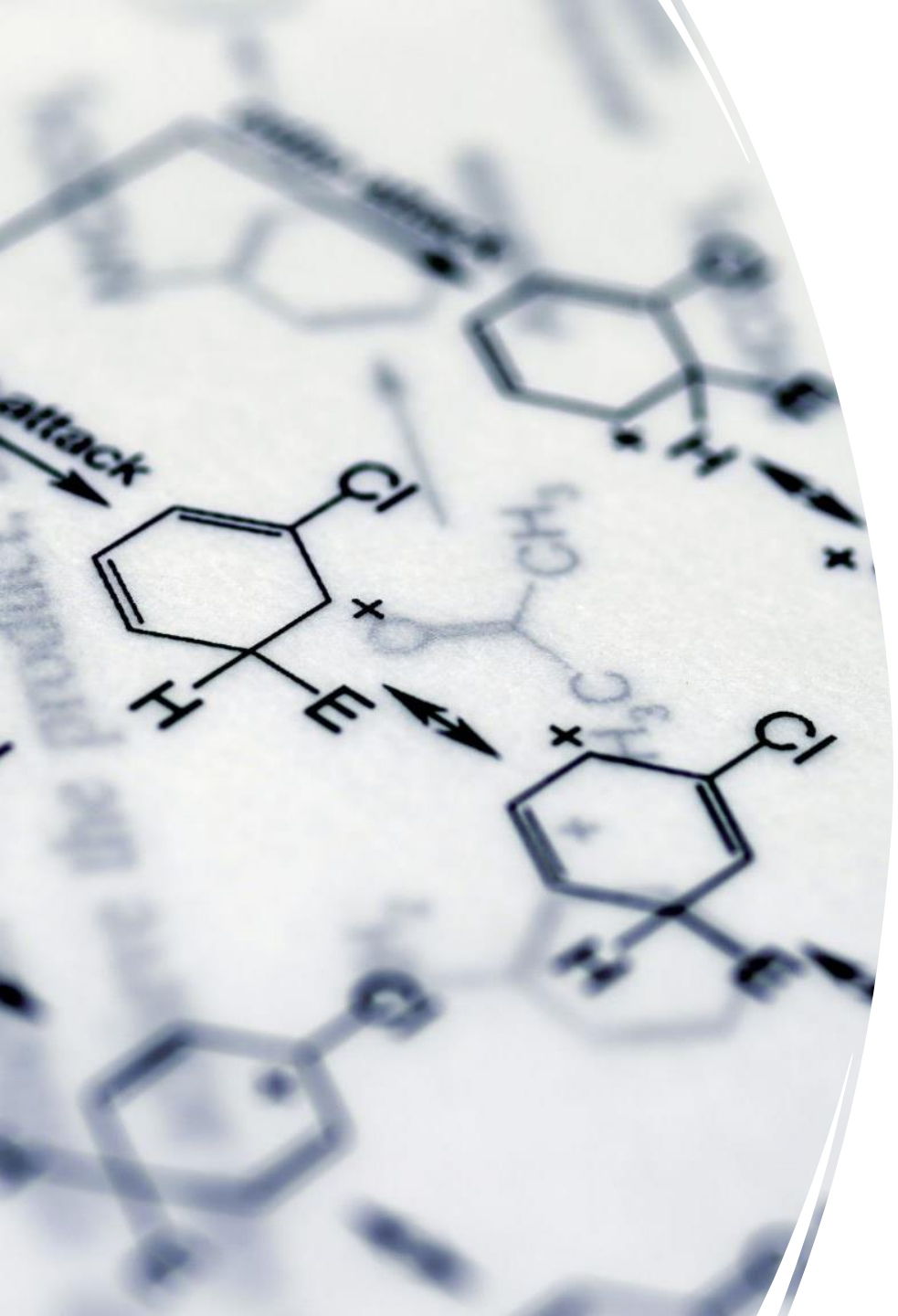
Expectations

- SIMCODES is a full time, 10-week research program.
  - NSF requires you participate in at least 8 weeks of the program.
  - You are expected to work full time (40 hours/week).
  - Hours are by default 9 AM – 5 PM but can be negotiated with your mentor.
- You must reside in the residence hall for the 10 weeks.
  - If your mentor allows you to work remotely, you must stay in Ames.
- You are strongly encouraged to attend extracurricular and professional development activities offered throughout the summer.
  - Events will be at https://simcodes-isu.github.io/calendar/.
  - We will also send reminders.
- You are expected to follow all federal, state, and local laws as well as all AMES and ISU policies.

Expectations Cont.

- You are expected to follow AMES and ISU's Information Technology policies.
  - Don't assume privacy.
  - Don't pirate movies/music.
  - Don't go to NSFW websites.
  - Don't mine cryptocurrency.
- Failure to meet these expectations may result in removal from SIMCODES.





Research Results

- Any intellectual property developed during the 10 weeks is the property of ISU.
  - In practice, this primarily prohibits you from patenting or otherwise trying to own/monetize the results.
- If you publish results, they should be acknowledged as:
  - This material is based upon work supported by the National Science Foundation under Grant No. 2348724. Any Opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

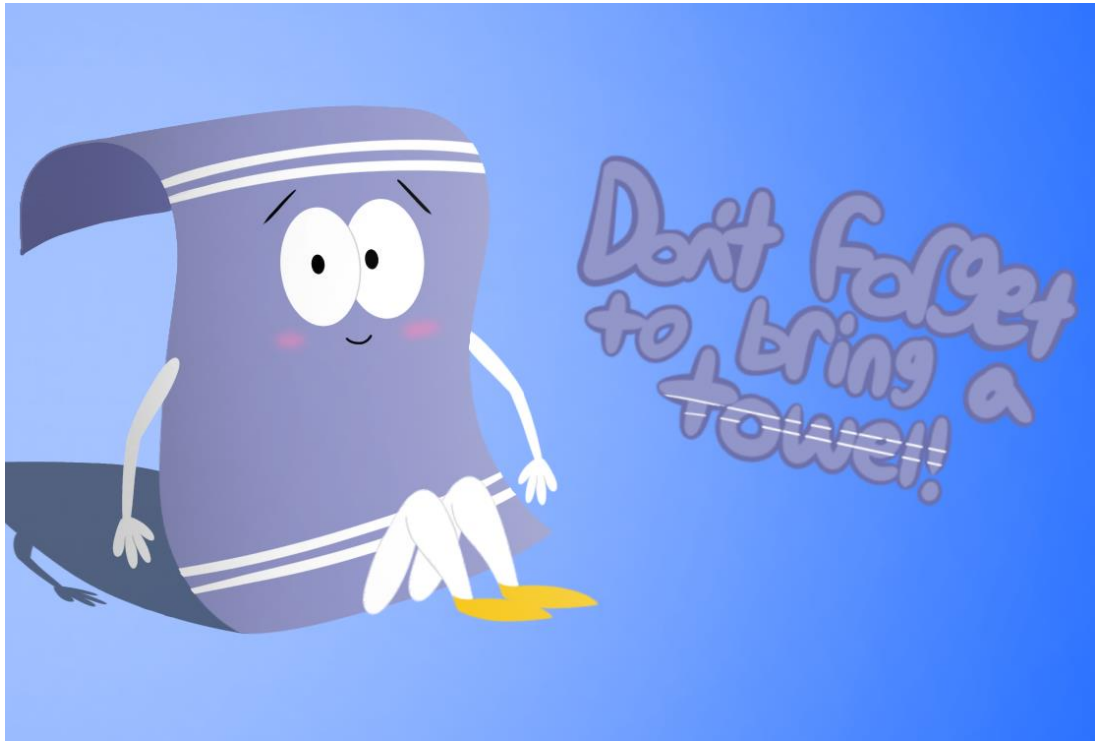
Code of Conduct

Code of Conduct

- We have a zero-tolerance policy for harassment and discrimination of any kind among any SIMCODES participants.
  - No, your first amendment rights do not protect harassment/discrimination in SIMCODES settings.
  - These policies also apply in the residency halls and at extracurricular activities.
- Be cognizant that others may come from different cultures and share different beliefs, i.e., we all have different thresholds and topics that we find offensive.
  - If you must think about whether something is offensive or not, assume it is.

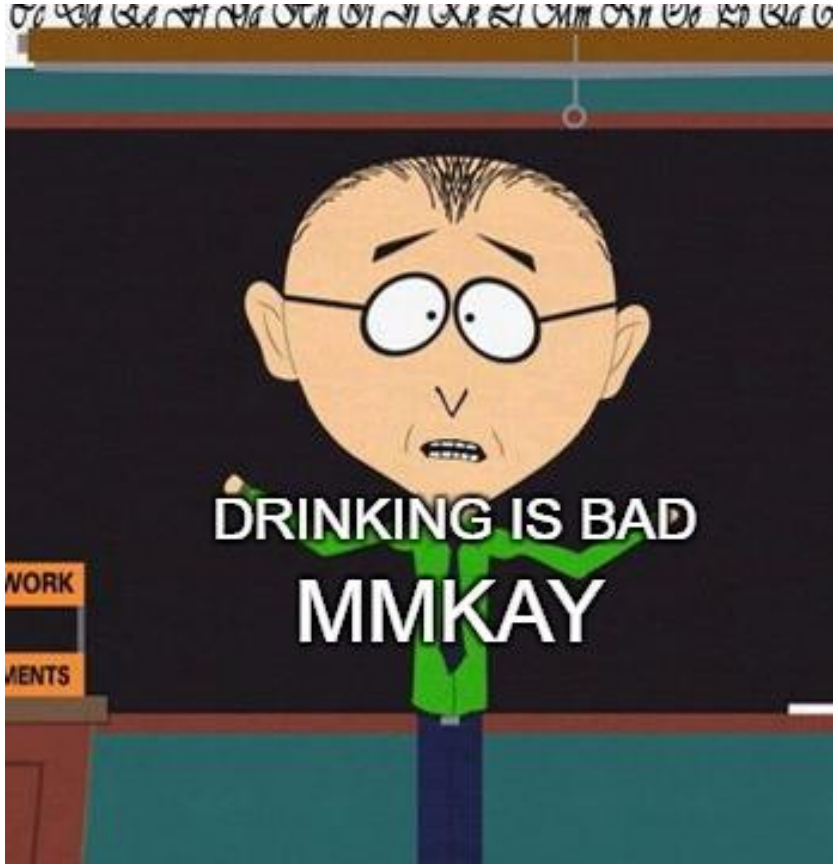
Notes on Smoking

- Smoking includes cigarettes, cigars, pipes, hookahs, e-cigarettes, e-pipe, e-hookah, vape pens, etc.
- ISU is a smoke-free campus. You may not smoke in any ISU building (including your apartment) or on any ISU owned property (including parking lots).
  - I have “no idea” where people go to smoke.
- Marijuana is illegal in Iowa for recreational use. Medical marijuana is legal. CBD is legal.
  - Some stores sell THC-infused beverages. My understanding is that this somehow skirts the law, but I am not a lawyer, and if you partake of these you do so at your own risk (I believe you have to be 21 to buy them).



Notes on Alcohol

- This is a college town with a college campus in it. Alcohol is plentiful and readily available.
- If you are legal drinking age (21 in Iowa) you may possess/consume alcohol in the apartments (full policy here), regardless of your roommates' ages.
- Some of the extracurricular activities may provide the opportunity to purchase alcohol. If you are of legal drinking you may partake if you want (you should not feel pressured to do so though).
  - SIMCODES can not reimburse alcohol purchases.
  - If you partake, please do so responsibly.
- Ames is a college town. The police readily hand out alcohol citations (by far and away underage drinking is the number one reason for arrest in Ames).



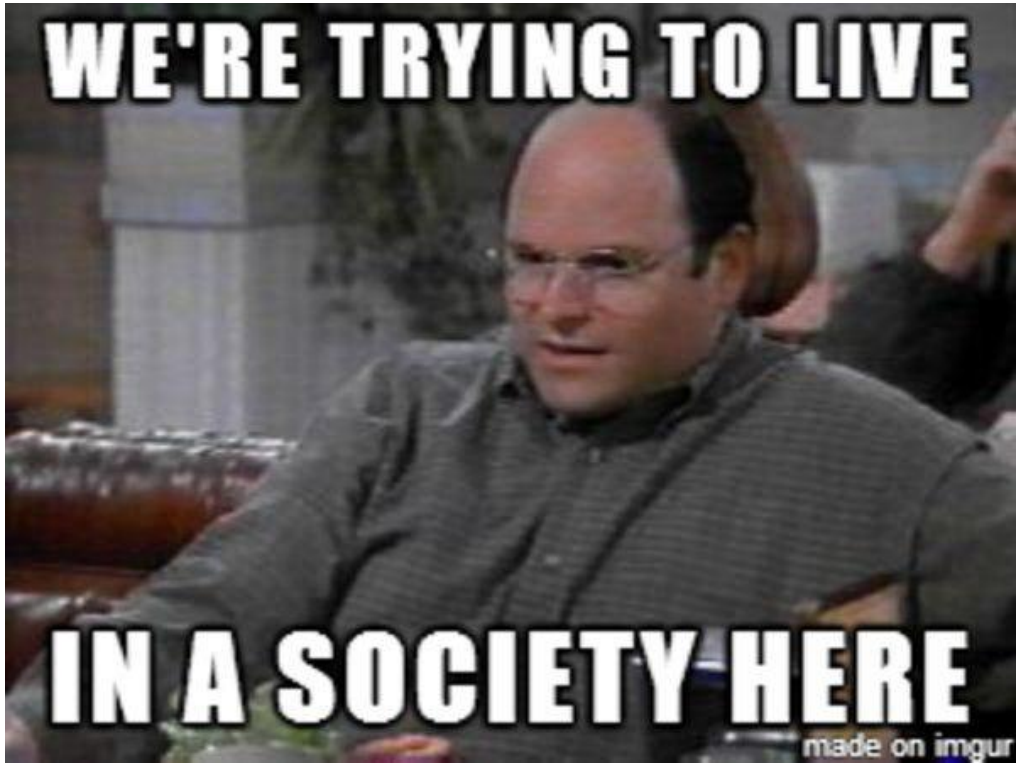
Notes on Other Drugs

- Caffeine is plentiful in Ames.
  - Courtyard Café is about two buildings away from AMES and the chemistry buildings.
  - If you are a Starbucks purest, closest one is on Lincoln Way by the Memorial Union.
  - Many departments and research spaces have coffee makers, check with your mentor if you're looking for free coffee.
- Unless you have a prescription, all other controlled substances are are illegal in Iowa.

Research Integrity

What is Research Integrity?

- Wikipedia: “The aspect of research ethics that deals with best practice or rules of professional practice of scientists.”
- Scientific “society” assumes that published, and peer-reviewed scientific results are credible.
  - “peer-review” is the process by which a study is scrutinized by experts in the field before it is published. In theory, this process catches errors in logic, procedure, or interpretation. In practice, things slip through.
- The rapid rate of scientific advancement is predicated on “I shouldn’t have to check your work (unless I want to).”

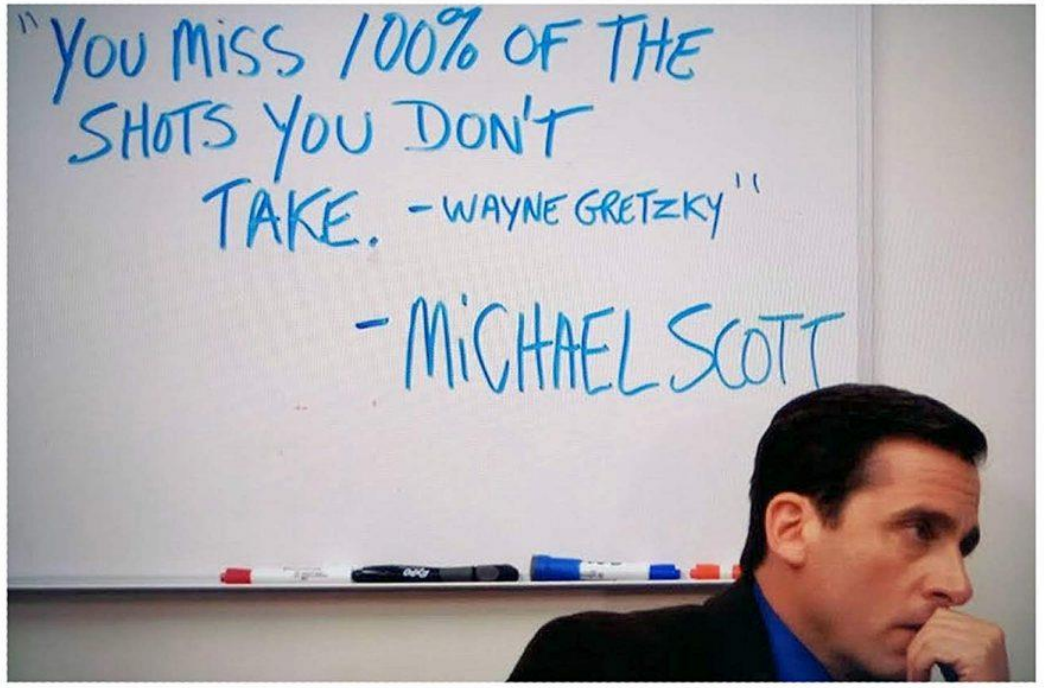


Research Integrity Policies

- You will be doing research meant to contribute to the scientific community. In turn, the scientific community expects your contribution to ensure the integrity of research.
- Research integrity varies a bit among research fields, so check with your mentor to make sure these slides cover your research.
- Failure to conduct research ethically may be grounds for immediate dismissal from SIMCODES.
    - Will depend on the severity. We're not going to throw you out because you didn't credit where you took a meme from (like someone around here…).

Citations and Attributions

- If someone else had an idea before you, you must cite it.
  - You are expected to perform due diligence to see if the idea exists.
- If you use someone else's data, software, images, or words, you must cite/attribute it.
- Different fields have different citation/attribution styles. Check with your mentor for your field's requirements.
- In general, I recommend noting:
  - The digital object identifier (DOI), if it has one. This will help you go back and get details (authors, title, etc.) later.
  - The URL and the date for websites.
  - The version for software you use.

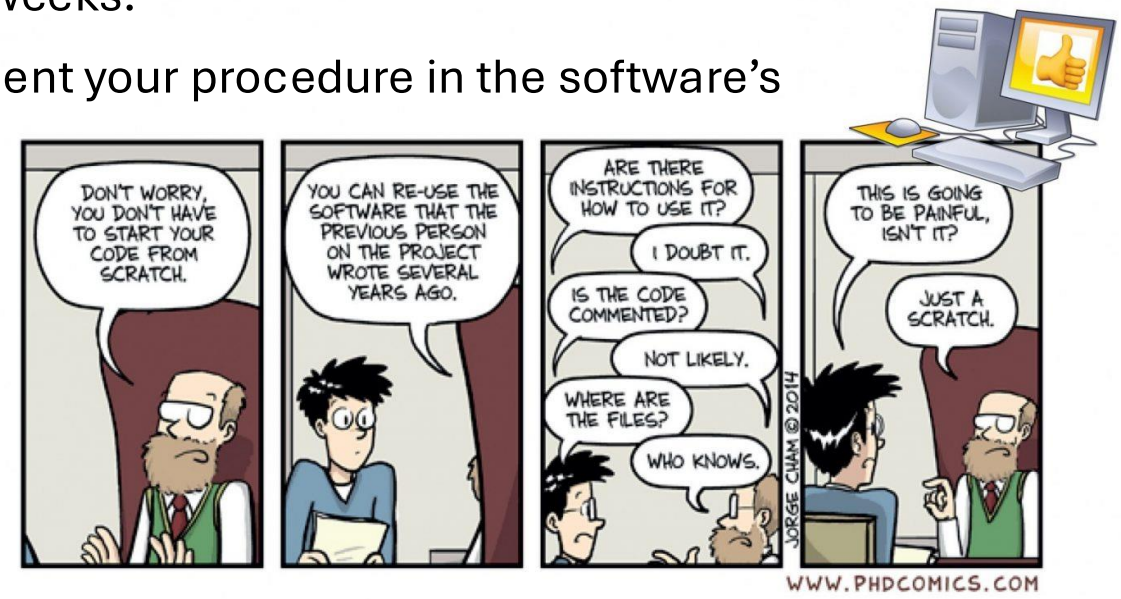


Data Transparency

- Data includes, but is not limited to numerical results, figures, software, posters, and presentations.
- You should save all research data, including undesirable results.
  - Examples of “undesirable” results are calculations that failed to converge and data that contradicts your hypotheses,
  - This does not refer to results from testing/debugging software unless the process is to be published.
  - Storage is cheap. If you’re not sure if you’ll need it, save it!
- It is your responsibility to collect “metadata” along with data.
  - Metadata = data about data.
  - Metadata is usually research specific, so check with your mentor.
- All data is subject to SIMCODES’s data policy (covered later).

Replicability

- Scientific results need to be reproducible.
- To ensure reproducibility it is important to document the procedure used to compute results.
  - In the moment it always feels like “I’ll remember how I obtained this result later.”
  - In practice, you’ll likely forget key details in two to three weeks.
- For software projects it is strongly recommended you document your procedure in the software’s documentation.
  - Provides the next person to use your software guidance.
  - Easy to find/cite later.

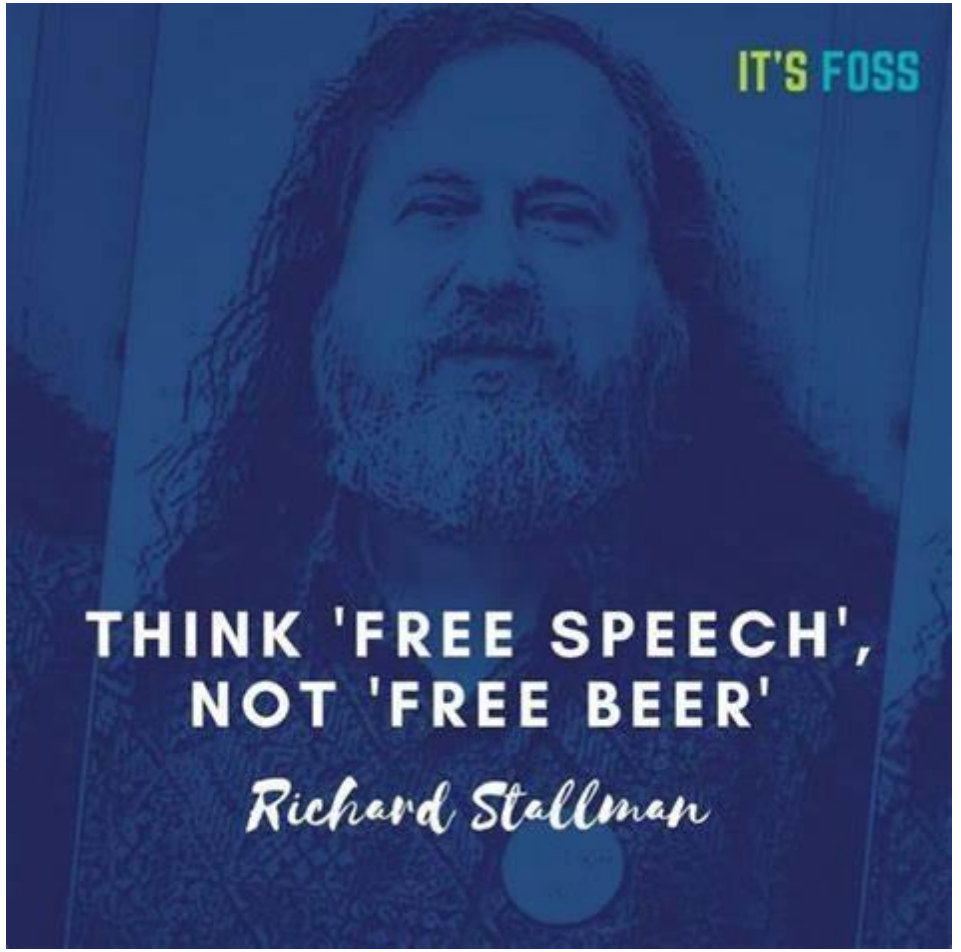


SIMCODES Data Policy

- Most funding agencies require data management plans. These are formal writeups of your data policies.
  - Partly to ensure scientific integrity.
  - Partly to ensure that data paid for with taxpayer dollars is accessible by the public.
  - Partly to ensure data is machine-readable for potential future AI/ML research.
- SIMCODES data policies provide specificity to the abstract points we just discussed.
- All members of SIMCODES are obligated to follow the same data policies.
- Future tutorials will provide you the training needed to comply with these policies.

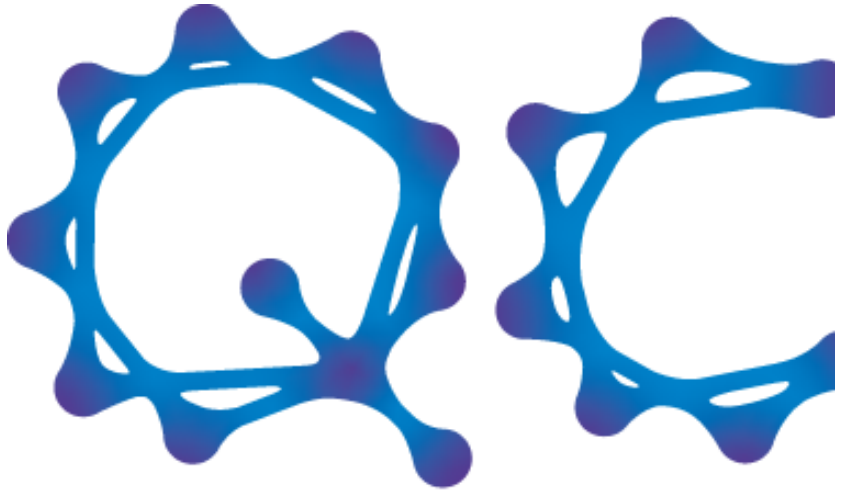
Data Access

- All data created (including software) during SIMCODES will be made publicly available and will be distributed as free and open-source.
  - Publicly available means anyone with an internet connection should be able to get your data.
  - Open-source means anyone with an internet connection can get the source code for the software.
- All SIMCODES projects resulting in new software will be developed as GitHub repositories owned by the SIMCODES-ISU organization.
- If you make contributions to existing software, you must have a plan in place to make those contributions public and open-source.
- Exception: public release of data may be embargoed until results are published.



Data Formats

- SIMCODES recommends all software development be done in Python.
  - Exception: you are contributing to an existing software package written in a different language.
- SIMCODES has agreed to store data in the formats created by MolSSI's QCArchive when applicable.
  - Applies primarily to inputs/outputs to/from computational chemistry packages.
  - AI/ML models should rely on these formats for input/output when possible.
- Presentations and posters will be stored in PDF format (and optionally in native formats).
- Storage formats of AI/ML models will be left up to the mentors.

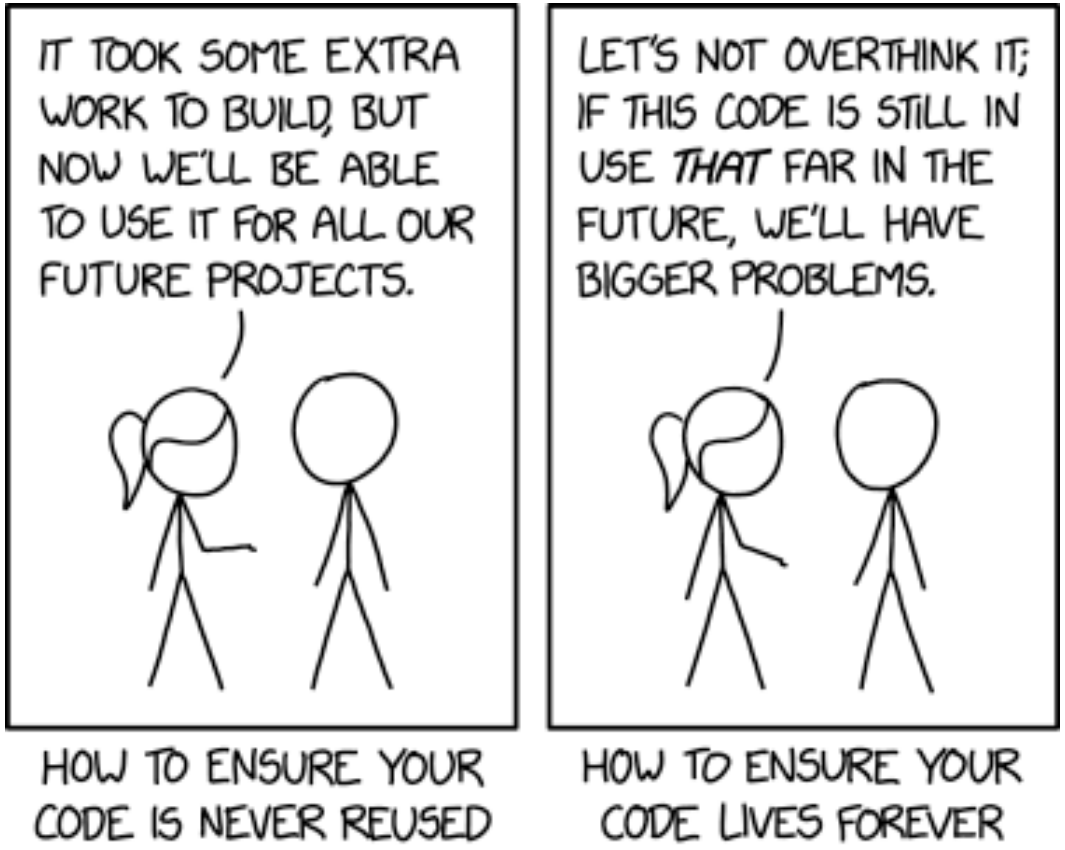


QCArchive

A MolSSI Projec

Data Reuse, Redistribution, and Archival

- Data generated during SIMCODES must be licensed and distributed under an open-source license (we recommend Apache 2.0).
  - Exception: contributions to existing software are done subject to the existing license.
- Data generated during SIMCODES will remain on GitHub indefinitely.
  - If you are done with your project, and there is no plan to continue development, it may be archived.
  - Archived projects are still publicly visible they just can't be extended.
- You are welcome to leverage your project (or any SIMCODES project) in future research.



Getting Help

Grievance Policy

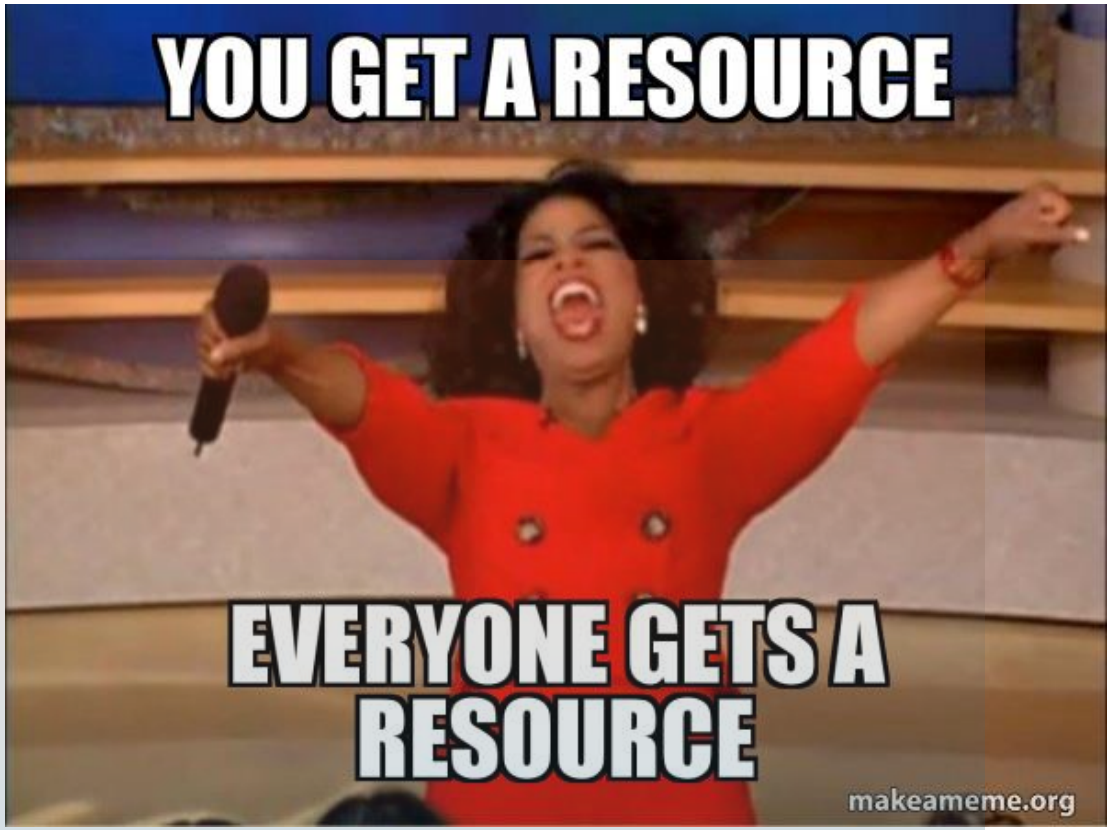
1. Informal resolution. Participants are encouraged to try to resolve conflicts directly or with their mentors.
2. Grievance submission. If the problem is not resolved from step 1 submit a written grievance to Grievance Officer (Ryan).
   - Make sure you include all parties involved, details, and evidence.
3. Review & investigation. The Grievance Officer with confidentially review the complaint and offer a resolution within 10 business days.
4. Appeal. If dissatisfied, the complaint can be elevated to the Principle Investigator (Theresa).

- Retaliation against complainants is strictly prohibited.

Research Questions

- Success in SIMCODES requires rapid progress.
  - Barriers naturally arise during research.
  - Can not afford to be stuck for long.
- Your mentor and cohort members should be your lifeline.
  - Make sure you are engaging regularly.
  - Slack is a convenient option for general questions since you know everyone has access.
  - GitHub is convenient for questions pertaining to specifics of software hosted on GitHub.
  - Ultimately, use what works for you, your mentor, and the rest of the cohort.

Resources



- For help writing software:
  - Python tutorials: MolSSI
  - Best practices: MolSSI
  - General help: Stack Overflow https://stackoverflow.com
  - Deep dives: GeeksForGeeks https://www.geeksforgeeks.org
- For help understanding computational chemistry:
  - Introduction: MolSSI
  - Deep dive: Psi4
- For help with AI/ML:
  - Introduction: MolSSI
  - Deep dive: Deep Learning for Molecules & Materials

Acknowledgements

- PowerPoint’s Designer for making the slides look pretty.
- The internet for memes.
- NSF for funding SIMCODES