

CompE5: End User Analysis

Snowmass Parallel



Amy Roberts (UC Denver, CDMS)

Peter Onyisi (UT Austin, ATLAS)

Gavin S. Davies (U. Mississippi, NOvA/DUNE/EMPHATIC)

<https://snowmass21.org/computational/analysis>



Text Editor

- Vim
- Or Emacs?
- If you don't know what these are, five points to Gryffindor and google "text editor wars"

ROOT

- List one thing you appreciate about ROOT
- You can do it!
- I believe in you!





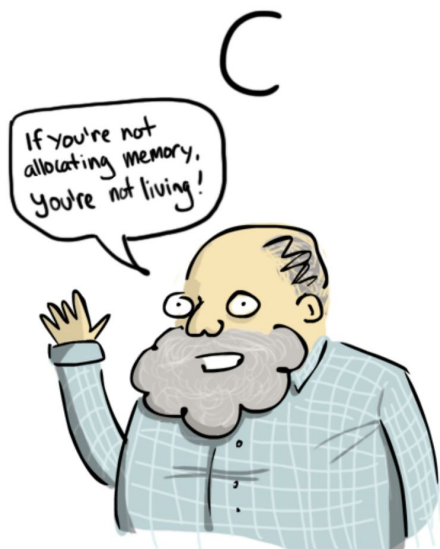
Plotting

- Sketch a plot you've made that you love
- Or a plot that you made a student make that you love #lolsob
- What did you use to make the plot?



Programming

- What programming language(s) are you excited about?



Or something else?



Discussion notes: languages

- We don't want to be locked into Python/C++ forever - new languages bring new capabilities! How do we support new languages while maintaining maintainability?
 - Lindsey: FFI layer. We have a strong need for interoperability and supportability. Support these might also provide support for language evolution.
 - Jerry: Julia for HEP analysis have made significant progress between 2020-2022, should mention it?
 - Ankur: Julia has been really nice for migrating people off Matlab and providing easy exposure to optimizing code and scaling up computational scale
 - Amy: does LLVM help us here?
 - Ian: The driving force behind language transitions isn't us! We may not have a ton of influence/control!
 - Axel: We can participate/give feedback to, e.g., hardware vendor language extensions and the ISO C++ committee (as ROOT does), and thus can influence industry trends. We should continue to integrate with these ecosystems for our needs & extend them.

Discussion: Training

- Is easier the right approach?
 - Lindsey: easy to read/debug is paramount
- Alex: People's interest in computing is varied.
 - Amy: I think we're discussing end-user analysis expectations/requirements, support for software development needs in the community, and training. End-user analysis tools don't have to address all these
- Doa:
- "What is the set of computing skills analyzer need?"
 - Rob: No clear consensus! Involved in a small experiment and recently lost an entire computing cadre. One of the issues is they've relied on magic postdocs instead of building a community. Amy: THIS IS ALL OF US. Rob: we really do need general skills.
 - Ian: When we create end-user analysis tools, how black-box is this? Is it a technical activity? Do we require people to have more general training?
 - Gabriel Perdue: it would be ideal to incorporate this into the curriculum
 - Doa: I took some of this as an undergraduate but really needed to learn more for

Discussion: Analysis preservation

- Amy: what do we do with non-structured metadata like wiki pages, email chains?
 - Giordon: If you are properly structuring your analysis within GitHub, etc. then nonstructured metadata is less of an issue. Part of the issue is a gap between interest in learning these tools and using these tools (like Docker containers). We need education and training and awareness. There is also a validation hurdle - having your analysis code/workflow in a github repo is not enough, you need someone to try to test it. A recommendation needs to include support for both these aspects
 - Rob: I've had trouble finding a funding path for this. Labs provide services for experiments and expect experiments to do this. Experiments expect labs to do this!
 - Peter: this is the kind of thing we can make recommendations on
 - Jeremy: How do we make the argument that this, out of all the other emergencies, needs funding
 - Ian the Blasphemer: We've struggled/failed with this for the last 70 years. Outside experiments it's virtually impossible to reproduce experiments. Why do we always fail at this? Is it really

General comments about the report

- Heidi: important to talk about exactly what the HSF is doing, they've done a lot of work and it's important to know what that is
- Daniel: when will a draft be available?

- Alex: should we think about the transition from analysis code as a small set of libraries/scripts to something that's maintained by someone (sometimes a lab)?
- Heidi: the first analysis framework is often not the best. We went with “let a thousand flowers bloom,” sometimes there's real value in letting people try things and have people vote with their feet.
 - Alex: yes, this can also become a maintenance problem
- Heidi: end analysis and simulation data are two entirely different sets of need, you can't use the same frameworks/code for both of these. Alex: cannot agree enough!
- Giordan: On lab-funded vs. university-funded code. Major features are now being dropped because the university can't afford to maintain the code anymore. From a preservation perspective, how do we preserve code that's not in our scope?