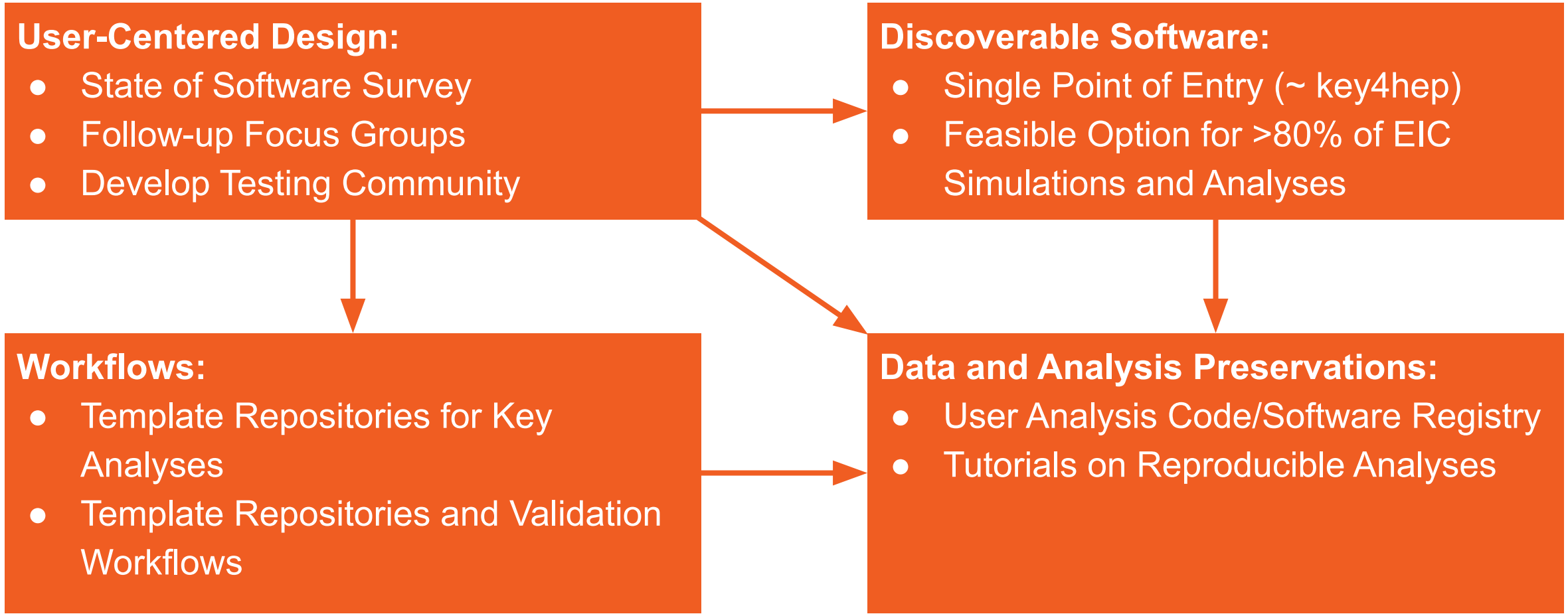
ELECTRON ION COLLIDER USER GROUP
USER SURVEY AND FOCUS GROUPS
First Insights

1. State of Software Survey: First annual survey of software use in the EIC User Group.

2. User Focus Groups: In-depth follow-up discussions, at first based on career stage.

by Dave Chopard (JLab), Wouter Deconinck (Manitoba), Markus Diefenthaler (JLab), Rebecca Duckett (JLab), Sylvester Joosten (ANL), and Kolja Kauder (BNL).

User-Centered Design: Listen to the Users, and/then Develop Software



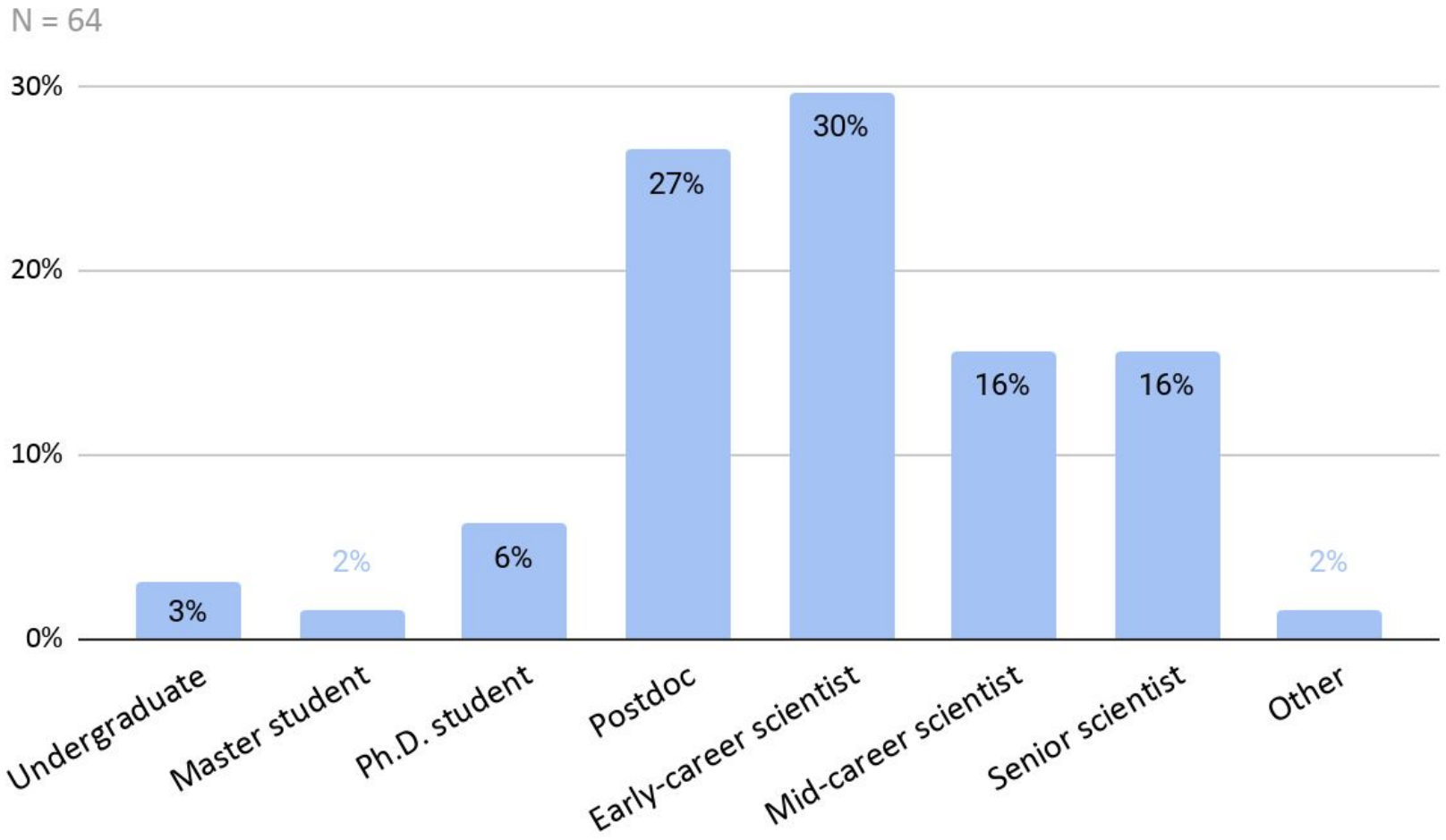
ELECTRON ION COLLIDER USER GROUP

1. STATE OF SOFTWARE SURVEY

Notes:

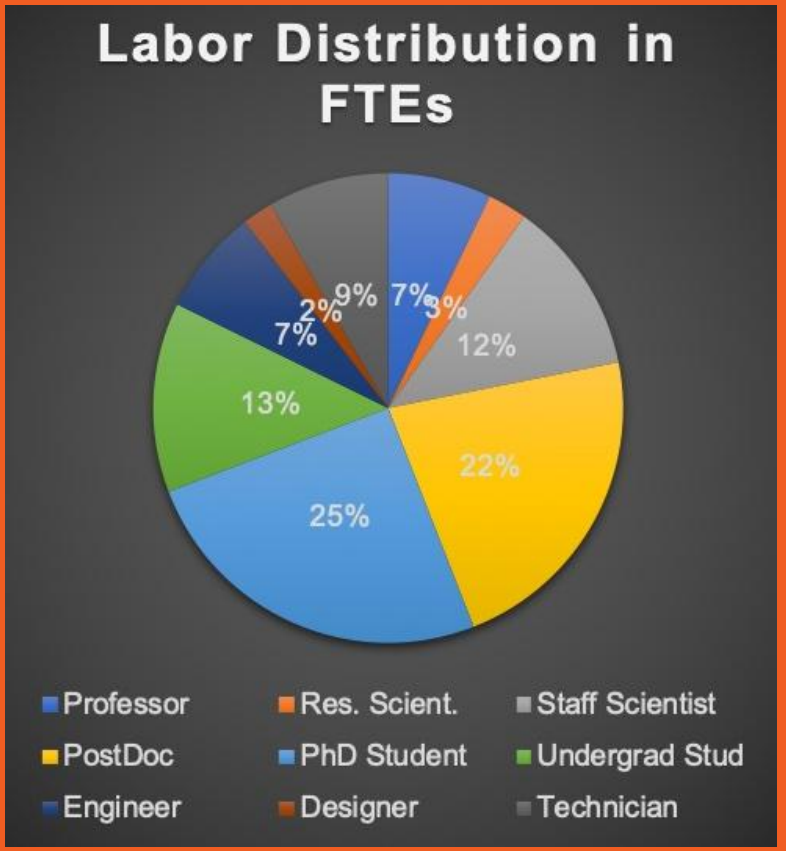
- Distribution methodology: Emails and reminders to EICUG mailing list.
- This year’s data collection has been completed. Your suggestions may be implemented in the next state of software survey (schedule for early 2022).
- A careful balance between collecting detailed information and avoiding survey fatigue means that some tools could not be include as predetermined options.

What is your current role in the EIC project?

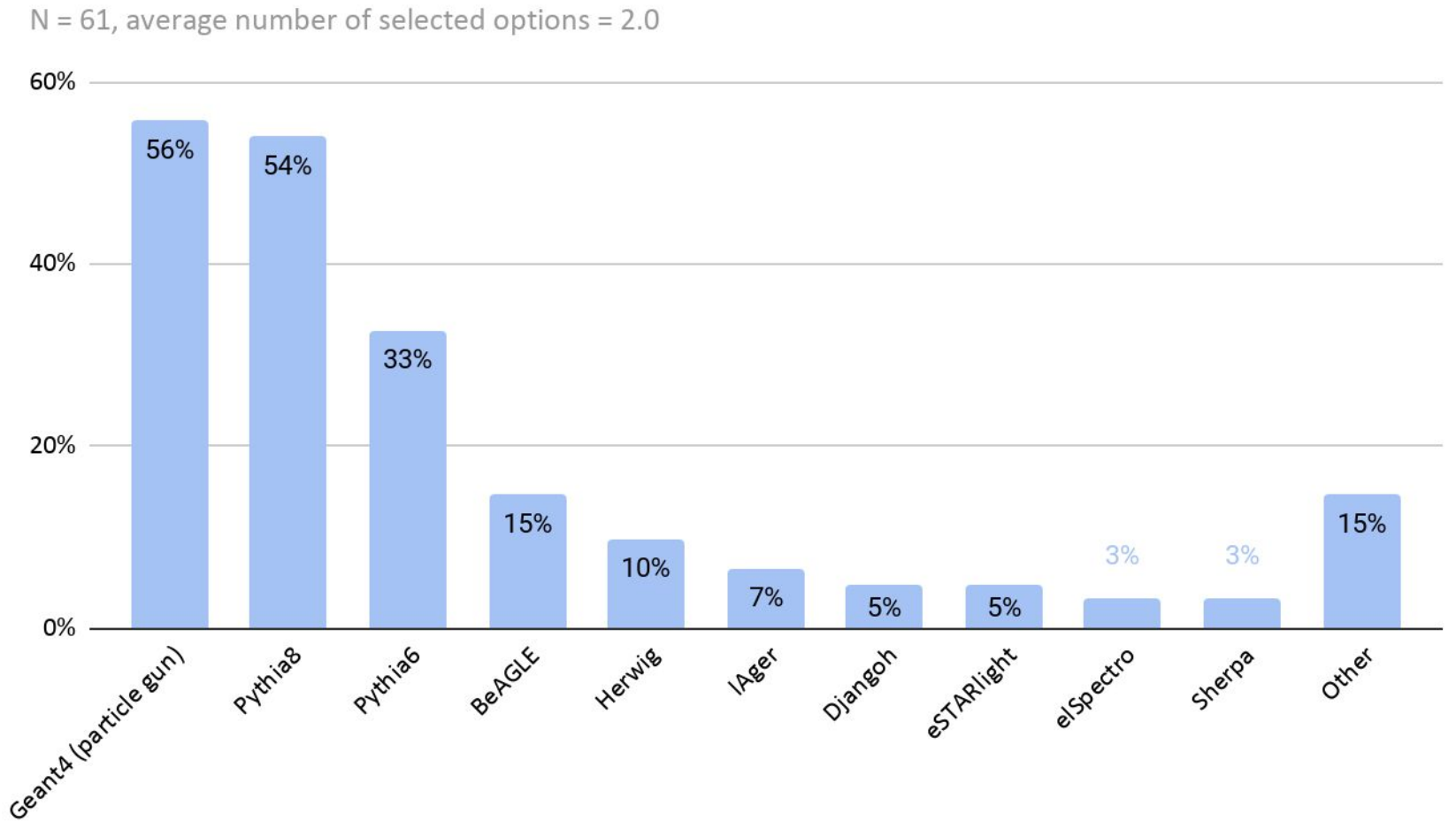


Other (N = 1): semi-retired senior researcher

Feedback from Expression of Interests (link)
Contributions from Ph.D. students will increase over time.

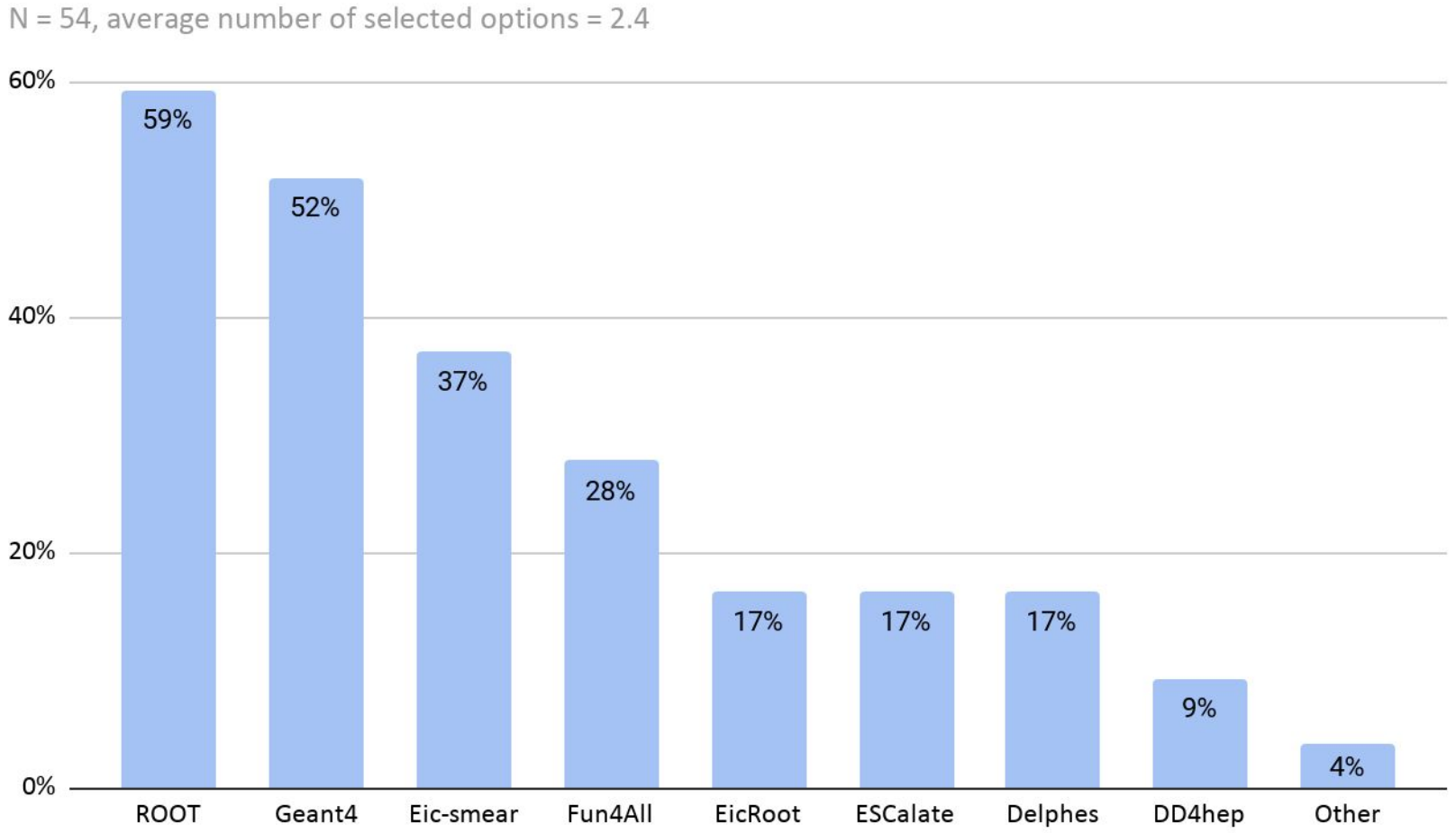


Over the past year, which physics event generation tools did you use for EIC simulations?



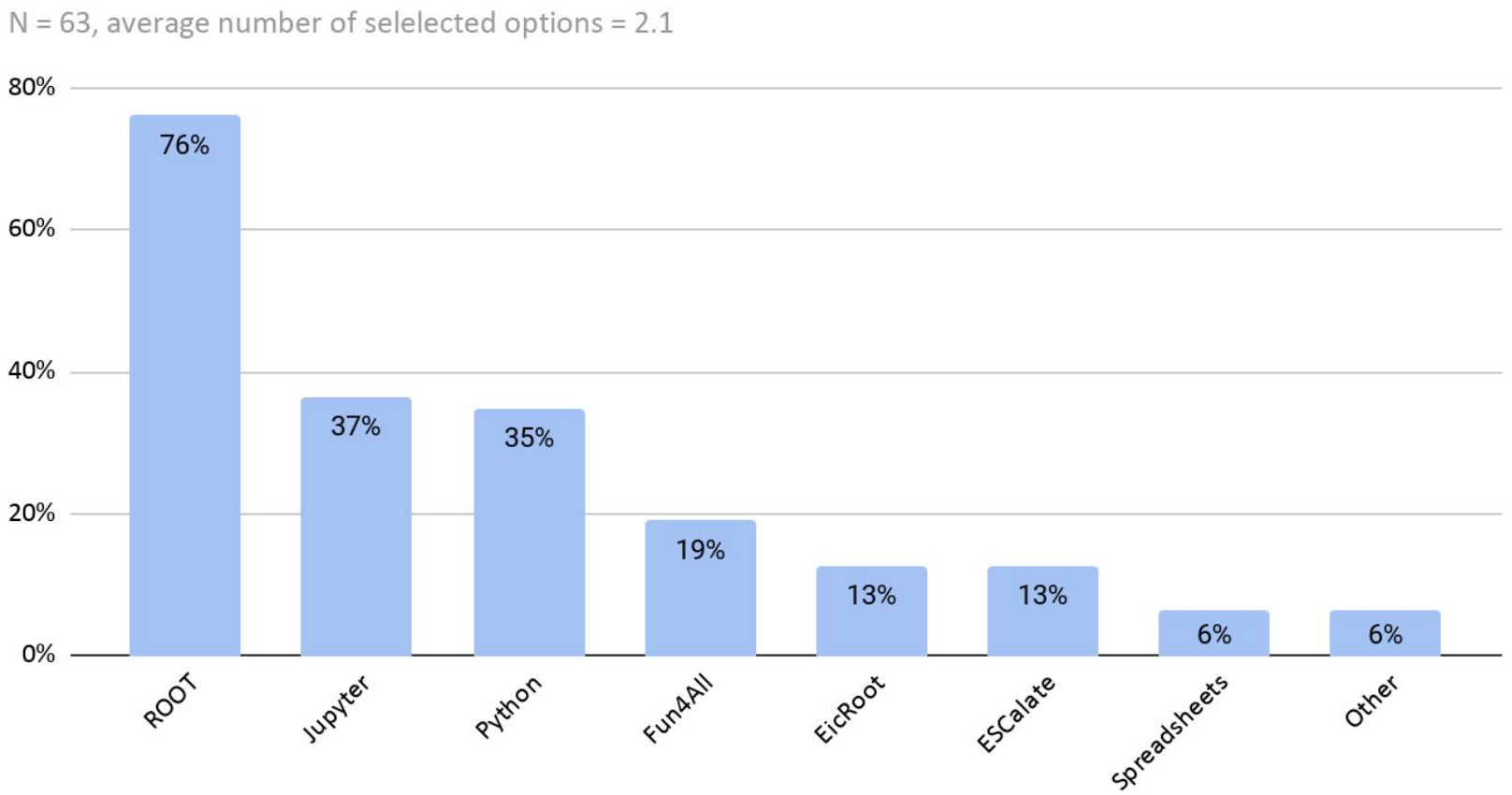
Other (N = 9): personal computer codes (N = 2), ACT, CLASDIS, ComptonRad, GRAPE-DILEPTON, MADX, MILOU, OPERA, RAYTRACE, Sartre, Topeg, ZGOUBI

Over the past year, which detector simulation tools did you use for EIC simulations?



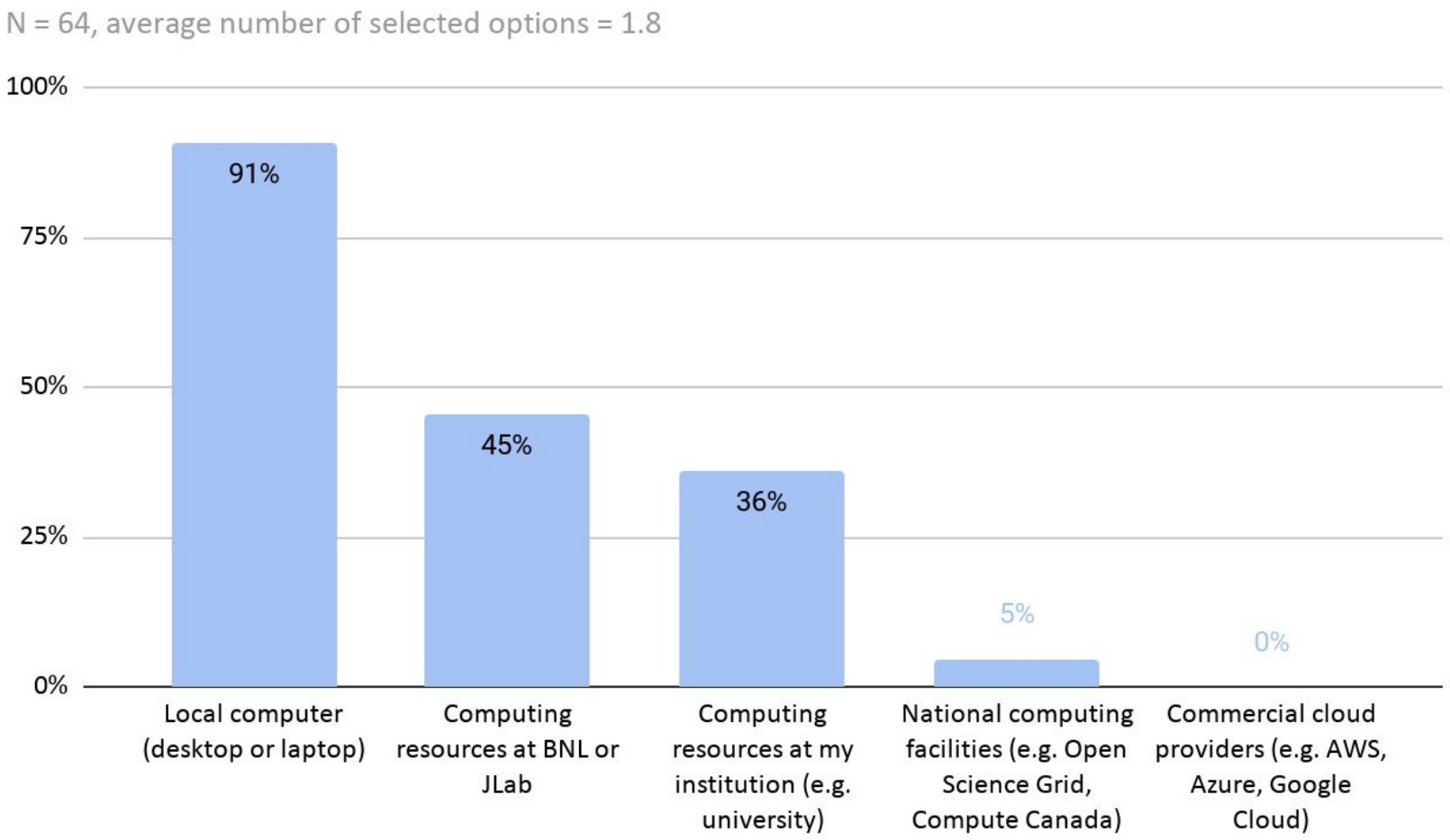
Other (N = 2): GEMC, RAYTRACE

Over the past year, which analysis tool(s) did you use for EIC simulations?

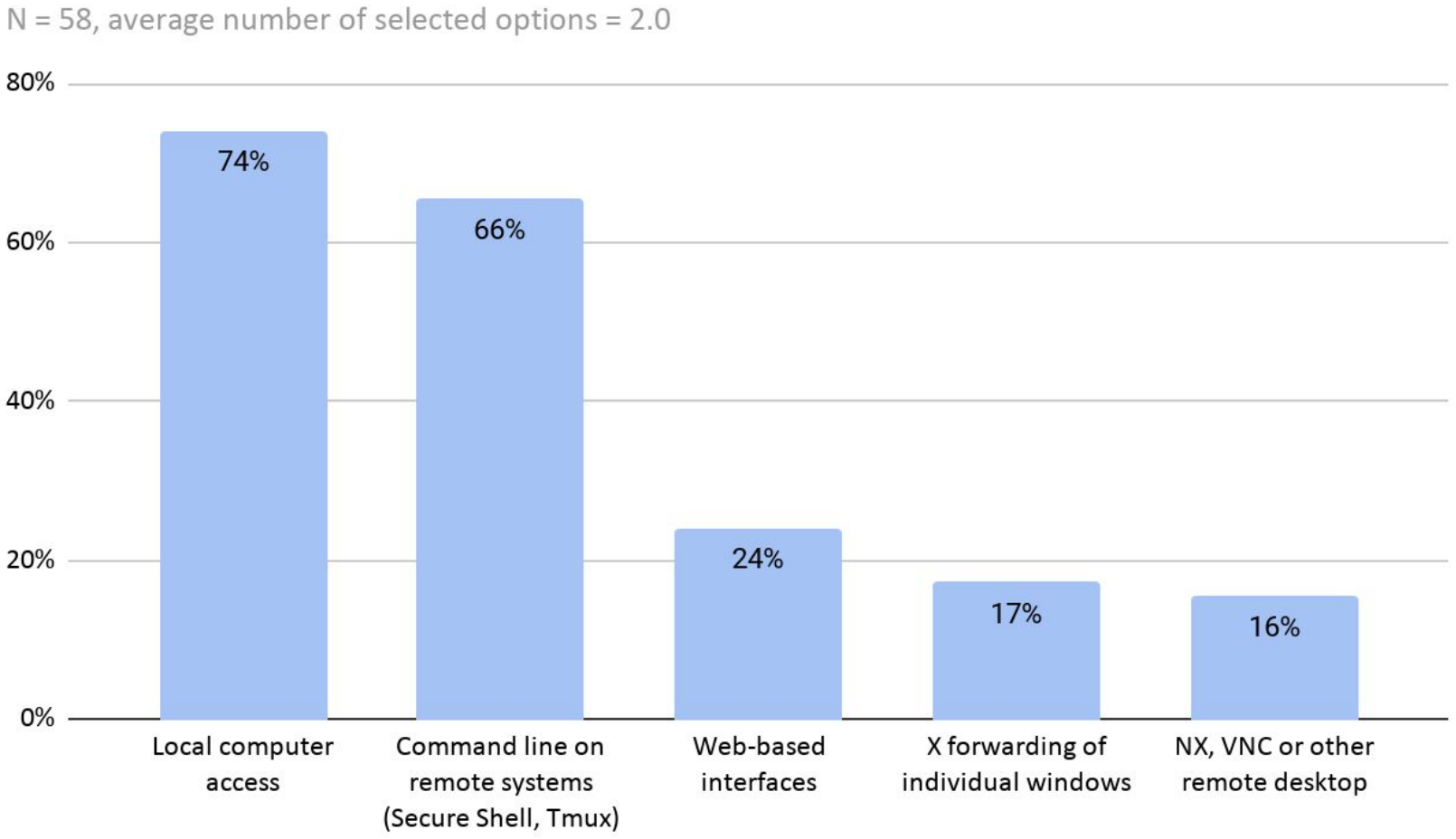


Other (N = 4): Rivet, ACE3P, jas4pp, custom codes

Over the past year, which resources did you use for EIC simulation and analysis?



Over the past year, how did you access the computing resources for EIC analysis?



Do you have any comments on your current experience with EIC Software?

N = 9

There are too many generators and simulation tools used at the moment.

5 x

- Lack of documentation.
- More tutorials would be beneficial.

3 x

The group should focus on full Geant4 simulation.

1 x

ELECTRON ION COLLIDER USER GROUP

2. USER FOCUS GROUPS

Notes:

- Selection methodology: Volunteered during focus groups or personal contacts.
- This year’s data collection based on career stage has nearly been completed.
- Future focus groups may be targeted based on software survey responses.

First Round of Focus Groups

Grouping Criterion: Career Stage

- Graduate students (n = 2)
- Junior postdocs (n = 4)
- Senior postdocs (n = 3)
- Staff scientists (n = 5)
- Professors (n = 5, to be scheduled)
- Physicists in industry (n = 10, to be scheduled)

Approach: Prompted Discussion, 1 hour or longer

- What software are you currently using, and why?
- Are you able and do you feel comfortable performing the computational aspects?
- What software or computing barriers do you encounter in your research?
- What do you think stopping physicists from participating in simulation or analysis?
- Which software tools do you particularly enjoy using, even if not EIC related?

Towards Personas and User Profiles: Attributes

Attributes of users:

- Low vs. high experience with physics computing
- Self-identification as user vs. developer
- Desire for guidance vs. self-starter mentality
- Need for custom software vs. availability of off-the-shelf functionality
- Positive vs. negative attitudes towards the process of writing software
- Positive vs. negative attitudes towards other users of the software
- Emotional or career investment in software
- Low vs. high ability to influence community, through positional power or power of expertise

Connection with general personality traits:

- Openness to the risk of new experiences: conservative vs. creative
- Conscientiousness: quick-and-dirty hack jobs vs. elaborate architectures
- Ability to compromise: autonomous vs. cooperative development

Towards Personas and User Profiles: Methodology

We score all participants on each attribute, normalize, then use k-means clustering and principal component analysis to identify the groups and distinguishing directions in attribute space.

OR

We follow a qualitative approach where similar statements by multiple participants and recurring quotes are foregrounded, and where we attempt to identify common themes across focus groups.

?

Towards Personas and User Profiles

Along one set of axes (~experience, ~attitude towards writing software):

- Starting Scientist
  - new to the field, don't know how to get information, dependent on others
- Starting Scientist with CS/CE Experience
  - new to the field, but with programming experience, dependent on others
- Software Using Scientist I
  - not interested in programming but competent, wishes for more documentation
- Software Using Scientist II
  - likes programming, contributes documentation to projects
- Software Developing Scientist I
  - active developers of large projects or frameworks
- Software Developing Scientist II
  - high-level perspective from experience
- Software Project Owning Scientist
  - in charge of an entire software project

Towards Personas and User Profiles

Along another set of axes (~attitudes towards users):

- Software as a necessary tool: “I like software, to the extent that it helps me get physics done. Give me a good example, and I’ll use your software.”
- Software is not my strong suit: “I am a bad programmer. I know. I write crappy code. Don’t force me to share my code with others. I am ashamed of it.”
- Software as part of my research: “I use software tools for my research project. I feel comfortable in using the software and modifying it for my needs. I share my modifications but software development is not my priority.”
- Software is a social activity: “I like to write software with and for others. I know I can write software pretty well, and I want to help people who don’t like it.”
- Software emperors: “I write the best software. I know how physics software should be written. Just follow my software-imposed rules.”

Towards Personas and User Profiles: Comments

There is still a reluctance to having analysis live in jupyter notebooks.

Don’t tell me to sign up for a mailing list: send an invite around that I can click on.

I don’t use git. No one is going to use my code anyway.

Maybe we need dedicated people who only write documentation.

Being in multiple experiments makes it hard to get proficient in the software tools used in each.

Installing software can be frustrating due to obscure dependencies.

Tutorials seem aimed at showing off software, not really teaching how to do something.

Lost lot of time due to lack of documentation, out-of-date documentation.

Debugging code in containers often was laborious.

More than 70% of the time I have to go through the source code to figure it out.

If I am stuck, I send an email or ask on Mattermost.

I don’t like to bother the main developer and prefer asking a peer first.

ELECTRON ION COLLIDER USER GROUP
USER SURVEY AND FOCUS GROUPS

Thank you to everyone who participated in the EICUG State of Software survey and Focus Groups. The Software Working Group will repeat the survey at the end of 2021 to compare results as we continue to design and build the Electron-Ion Collider.

Next steps

- We will organize further focus group discussions that will result in personas and user stories.
- The user stories will provide input to software developers as to which users they are writing software for.

Appendix: Survey Questions

Q1. What is your current role in the EIC project?

- ○ Undergraduate student
- ○ Graduate student (M.Sc.)
- ○ Graduate student (Ph.D.)
- ○ Postdoctoral researcher
- ○ Early-career scientist (pre-tenure, assistant professor, staff scientist < 5 years)
- ○ Mid-career scientist (tenure, associate professor, staff scientist 5-10 years)
- ○ Senior scientist (full professor, staff scientist > 10 years)
- ○ Other (please specify)

  _____

Q2. Over the past year, which physics event generation tool(s) did you use for EIC simulations? Check all that apply.

- □ Geant4 (particle gun)
- □ BeAGLE
- □ Djangoh
- □ eSTARlight
- □ Herwig
- □ lAger
- □ Pythia8
- □ Pythia6
- □ Sartre
- □ Other (please specify)

  _____

Q3. Over the past year, which detector simulation tool(s) did you use for EIC simulations? Check all that apply.

- □ ROOT
- □ Geant4
- □ DD4hep
- □ Delphes
- □ Eic-smear
- □ EicRoot
- □ ESCalate
- □ Fun4All
- □ Other (please specify)

Q4. Over the past year, which analysis tool(s) did you use for EIC simulations? Check all that apply.

- □ ROOT
- □ EicRoot
- □ ESCalate
- □ Fun4All
- □ Jupyter
- □ Python (NumPy/Pandas/...)
- □ Spreadsheets
- □ Other (please specify)

  _____

Q5. Over the past year, which resources did you use for EIC simulation and analysis? Check all that apply.

- □ Local computer (desktop or laptop)
- □ Computing resources at my institution (e.g. university)
- □ Computing resources at BNL or JLab
- □ National computing facilities (e.g. Open Science Grid, Compute Canada)
- □ Commercial cloud providers (e.g. AWS, Azure, Google Cloud)

  _____

Q6. Over the past year, how did you access the computing resources for EIC analysis? Check all that apply.

- □ Local computer access
- □ Command line on remote systems (Secure Shell, Tmux)
- □ X forwarding of individual windows
- □ NX, VNC or other remote desktop
- □ Web-based interfaces

  _____

Q7. Do you have any comments on your current experience with EIC Software?

  _____

Q8. Are you interested in volunteering for future focus group discussions on EIC Software? If so, please enter your email address.