SE4Science'21 Notes

The short link to this document is: http://bit.ly/SE4Science21Notes

Intro slides are available from: http://bit.ly/SE4Science21Slides

Attendees

Please add your name and organisation, along with your email if you would like to be added to the SE4Science mailing list.

- Neil Chue Hong, Software Sustainability Institute / University of Edinburgh, N.ChueHong@software.ac.uk
- Jeff Carver, University of Alabama, carver@cs.ua.edu
- Anna-Lena Lamprecht, Utrecht University, a.l.lamprecht@uu.nl
- Jeremy Cohen, Imperial College London, jeremy.cohen@imperial.ac.uk
- Matthew Bluteau, UK Atomic Energy Authority, matthew.bluteau@ukaea.uk
- Caroline Jay, Software Sustainability Institute / University of Manchester, caroline.jay@manchester.ac.uk
- Nan Niu, University of Cincinnati, nan.niu@uc.edu
- Zedong Peng, University of Cincinnati, pengzd@mail.uc.edu

Agenda

All times are Central European Summer Time (UTC+2)

13:20-15:00 SE4Science Session 1 (Chair: Anna-Lena Lamprecht)
- 13:20   *Zedong Peng*, Xuanyi Lin, Nan Niu and Omar I. Abdul-Aziz. **I/O Associations in Scientific Software: A Study of SWMM**
- 13:40   *Neil Chue Hong*, Jeremy Cohen and Caroline Jay. **Understanding Equity, Diversity and Inclusivity Challenges Within the Research Software Community**
- 14:00   Tamara Lopez, *Caroline Jay* and Helen Sharp. **How has the COVID-19 Pandemic affected working conditions for Research Software Engineers?**

16:30-18:10 SE4Science Session 2 (Chair: Neil Chue Hong)
- 16:30   Introduction to Speed Blogging and Topics (Neil Chue Hong)
- 16:40   Discussion and scoping of article
- 17:00   Article writing
- 18:00   Closing remarks (Jeff Carver)

Notes

Zedong Peng, Xuanyi Lin, Nan Niu and Omar I. Abdul-Aziz.
I/O Associations in Scientific Software: A Study of SWMM

Slides and video:
https://drive.google.com/drive/folders/1RQjWS6iduEdvwILeywIebC1vXwdxXcm9?usp=sharing

- Q: How do you get an idea of which are the "most important" I/O variables? Is this based on the variable with the highest associations with the largest number of other variables? +1
  - We offer a couple of perspectives here. One is based on the 'support' and 'confidence' scores of association rule mining from the user forum of the scientific software. The top-ranked I/O associations, that is, the association rules with the highest support and confidence, could be considered as 'important' due to the usage data (i.e., those I/O's are likely to be used more heavily by the end users of the software). The other view could be performing actual testing (or program analysis), that is, changing the input and then checking whether the output is affected at all. If focusing on a particular I/O pair reveals a software defect, then practically, this I/O association is 'important'.
- Q: As SWMM started in 1971, I wonder how much the legacy code that is in there contributes to the testing complications. Would it be "easier" (from a testing perspective) to reimplement the system from scratch?
  - Currently, SWMM use GitHub as perform with SWMM developer, base on one of our research (https://doi.org/10.1016/j.jocs.2021.101347), we found 2953 unit tests and 58 regression tests at SWMM GitHub (https://github.com/OpenWaterAnalytics/Stormwater-Management-Model). I believe the computing engine will be very complex and require a lot of expertise. But I think using unit testing to validate that each unit of the software code performs as expected will be helpful for reimplementing the system from scratch.

Neil Chue Hong, Jeremy Cohen and Caroline Jay.
Understanding Equity, Diversity and Inclusivity Challenges Within the Research Software Community
- Slides: https://doi.org/10.6084/m9.figshare.14787858
- Preprint: https://arxiv.org/abs/2104.01712
- Q: Did you consider comparing this to the self-perception of the RSE community? I remember from one of the first RSE conferences that I attended that a keynote speaker talked about how inclusive she experienced the RSE community, that she finally had found a community where she felt home etc. This had more to do with skills and backgrounds etc than gender/ethnicity/disabilities, but I just wonder if RSEs think they are very high on EDI standards?
  - We have anecdotal evidence that people in the RSE community consider RSE specific events to be more diverse and inclusive, however the wider

population is still not as representative as it could be. A hypothesis to be examined is that there are a number of high-profile initiatives and groups following best practice, but there is still work to be done in the more dispersed, fragmented areas in which RSEs work (and computational science is carried out).

- Q: I take the point about this likely not being a pipeline problem, but I also wonder if there are some pipeline issues hiding underneath the coarseness of the categories considered in the diagram? E.g. "Physics & Astronomy" is quite a large subject area and there is likely quite large variation of some of these characteristics within it, where some of the more "computational areas" might tend to lead into RSE positions more often, and there might be existing imbalances at this precursor stage. I am assuming the data just isn't at this precision level yet?
  - We looked a little into this hypothesis in the paper but we need to get access to more fine-grained data from the UK Higher Education Statistics Agency, and the data isn't quite at this level of detail in the International RSE Survey.
    - "Over half of RSEs have a first degree in Physics and Astronomy or Computer Science. In the UK, 17% of CS undergraduates were female compared with 41% of physical sciences undergraduates [16], suggesting that RSEs come from the "computational" subset of a subject. However, within computer science research within the UK just under 23% of academics and researchers are female [14]. This perhaps suggests that a larger percentage of female CS undergraduates move on to a research or faculty position in CS than male undergraduates, or that female CS researchers are moving into the field after undergraduate studies in other areas. However, it could also be attributed to other factors such as industry hiring trends."

Tamara Lopez, Caroline Jay, Michel Wermelinger and Helen Sharp.
How has the COVID-19 Pandemic affected working conditions for Research Software Engineers?
- Q: Have you done any longer term follow-ups?
- Q: connecting to the previous talk, did you collect any information about protected characteristics, and other factors like caring responsibilities? I would imagine these having quite an influence. The well known trend that women tended to suffer with productivity compared to men (at least in academia) has been widely reported.
  - Partially answered in your conclusions :)
- .Q: What lessons from your work do you think will carry forward after the pandemic?

Discussion Topics for Speed Blogs

- Socio-cultural challenges to software engineering for computational science
- Testing techniques for computational sciences
- Domain specific languages and frameworks for computational science codes

Click on topic to go to document you will be writing in.