Training Subcommittee Minutes – February 27, 2018

Attendees: Marc Spiegelman (Chair), Andreas Mueller, Ryan Abernathey, George Garrett, Haim Waisman, Barbara Rockenbach, CJ Wright, Kyle Mandli, Tian Zhang, Rob Lane, Victoria Hamilton (Staff), Marley Bauce (Staff)

To Do

- Victoria and Marley to assume responsibility to develop a list of departmental points of contact, and to compile a list of Directors of Graduate Study for all departments, who will be the default representatives.
- Marley and CJ to work together to develop questions for students.
- Marley to speak with Michelle Benson about emerging federal requirements for training career award recipients in reproducibility, thereafter sending information to the Subcommittee.

Reminder of Subcommittee Charge and Timeline

Marc opens the meeting by reviewing the general charge of this Training Subcommittee: To make a recommendation to SRCPAC and the RCEC about (a) whether there is a need for training complementary to—and possibly as a prerequisite to--normal course offerings for graduate students, and (b) if there are common needs across departments that could be served well at an institutional level. For example, by establishing a basic hierarchy for guiding graduate students into the basics of research computing: Python, R, Git, etc. The focus of today's meeting is to confirm that there is actually a departmental need for these services. Marc speaks to his two departments – APAM and DEES –as having this problem, but is unsure whether there is a problem of scale, and whether we should be doing anything as an institution. Anecdotally, the Subcommittee believes there is a problem, but we will need data on this issue. On May 24th, the Research Computing Executive Committee will meet and expect a report from this Training Subcommittee.

We currently have five departmental points of contact, whereas there are 18 departments; Victoria and Marley will work to build this list quickly. Tian suggests that we also focus on the social science departments; Victoria and Marley will first contact the Deans for Humanities and Social Sciences, then reach out to the relevant department chairs to request participation.

Departmental Expectations Survey

Marc, Victoria, and Marley developed a draft survey of questions to send to departmental points of contact; Marc asks the Subcommittee whether there are any edits to these existing questions.

Haim, speaking for the Department of Civil Engineering and Engineering Mechanics, says he encounters a range of skills among incoming graduate students, from people who have never done anything, to advanced learners. He says it would be nice to have a standard so that everyone is on the same page. There is a need in the Department to know about basic numerical elements and algorithms and numerical integration; they typically use MatLab.

Rob asks whether prior experience in computation leads to adequate knowledge; Haim responds that if they have done it previously, it is a plus, but not always.

In Statistics, Tian notes that faculty are expected to teach computing; they teach modelling, but thereafter students are expected to learn on their own. It is increasingly difficult for students to do this well and transparently. Additionally, some students will initially struggle in computing, and therefore will be adverse to

learning computational applications in the future. Tian suggests a boot camp and/or short courses to demonstrate the wealth of resources, and that with a bit of preparation that everyone can do data-intensive research.

The Subcommittee recognizes that we have one – and only one – opportunity to collect data from departments, which makes this effort critical in terms of language and strategy employed: If we do not adequately anticipate needs and the ultimate reporting structure, we cannot go back to departments asking for a second installment of information. The Subcommittee then asks the following conceptual questions about the survey design:

- What skills and tools do we want students to know about?
- Do individual departments have required pre-requisites for applying to their graduate program, and what are they?
- Can we spell-out examples of skills and tools, and allow departments to "check all that apply"?

The Subcommittee then clarifies that it is highly likely that departments (and their faculty points of contact) may have differing answers and perspectives than the graduate students themselves, and that any departmental survey should be supplemented by a *student* survey. CJ and Marley will collaborate to develop this corresponding survey, and Marley will ultimately distribute it across the Morningside graduate student body.

Andreas then distributed the draft departmental survey to all Subcommittee members; the Subcommittee was asked to make any and all edits to the questions by March 13th, after which time Marley will begin developing the survey platform.

Boot Camp on Reproducible Research

Tian alerts the subcommittee to David Shilane, who is a Data Science Institute adjunct professor seeking to develop a new course on research computing reproducibility and reporting, and focusing on the techniques used to analyze data. Tian asked him to condense his semester-long course into an accelerated boot camp, which would allow both faculty and students to participate. The lecture will be complemented by a lab portion, where mostly students will stay for the longer lab section. Tian has received buy-in from GSAS, as well as the Deans of Science and Social Science. Tian suggests that with leadership from DSI and additional buy-in from SEAS, this could lay down a solid foundation in research computing. Ryan asks whether this will be a boot camp or a course; Tian responds that it can be both but *only if there is a need for both*, which would be bolstered via support by the Libraries. Kyle says he would require his students to take such a course; the Subcommittee also notes that soon the NIH will require formalized training for all of its K award recipients, which potentially this course would satisfy. Marley will follow-up with the Subcommittee with more information about this emerging federal requirement.

Rob asks when this boot camp would happen. Tian responds that the Data Science Institute can organize and facilitate this in collaboration with GSAS, but we would need to find a way to finance the operation. CJ notes that SEAS is pushing on PDL (Professional Development Lessons), and that they will expand their programs to PhD students, and may be interested in this topic.