

Shared Research Computing Policy Advisory Committee – Spring 2015 Meeting
Thursday, April 30th, 2015

11:30 a.m. – 1:00 p.m., 523 Butler Library
Minutes Report

Attendance

Kathryn Johnston, Astronomy (Chair of SRCPAC)
Rajendra Bose, Zuckerman Mind Brain Behavior Institute (Chair of Intercampus Subcommittee)
Greg Bryan, Astronomy (Chair of Yeti Operating & Executive Committees)

Maneesha Aggarwal, CUIT
Alexander Antoniadis, CUIT
Alex Bergier, CUIT
Paul Blaer, Computer Science
Harmen Bussemaker, Biological Sciences
Rob Cartolano, Libraries & Information Sciences
Jennifer Chambers (for Brent Stockwell), Biological Sciences
Alan Crosswell, CUIT
George Garrett, CUIT
Andrew Gerber, Psychiatry
Victoria Hamilton, Office of the Executive Vice President for Research
Halayn Hescok, CUIT
Rob Lane, CUIT
Derrick Lim, Center for Computational Learning Systems
Gaspere LoDuca, CUIT
Chris Marianetti, Applied Physics & Applied Mathematics
Matt McCoy, Office of the Executive Vice President for Research
Amy Nurnberger, Libraries & Information Sciences
Jerry Ostriker, Astronomy
Mahdad Parsi, Lamont-Doherty Earth Observatory
Dali Plavsic, CUIT
Ingrid K. Richter, Psychology
Evan Riehl (for Bentley MacLeod), Economics
Bruno Scap, Data Sciences Institute/Biomedical Informatics
Jochen Weber, Psychology
Michael Weisner, CUIT
Rebecca Yohannes, Public Health
Tian Zheng, Statistics
Allen Zweben, Social Work

Summary

At the Spring 2015 SRCPAC Committee meeting, members:

- Endorsed the plan of a next call for purchase in September 2015, and recommended the formation of a Faculty Advisory Committee to guide the RFP process
- Recommended a policy to allocate capacity among:
 - Normal community procurement cycles;
 - Recruits; and
 - Proposals that incorporate substantial computational resources

Introduction

Chair Kathryn Johnston called the meeting to order at 11:30 a.m., distributed the meeting [agenda](#), and asked all meeting participants to introduce themselves.

Intercampus Subcommittee

Raj Bose, Chair of the Intercampus Subcommittee, began his presentation by mentioning that the Intercampus Subcommittee had been enhanced by an expanded membership as recommended at the Fall SRCPAC meeting. With the addition of Harmen Bussemaker, Professor of Biological Sciences, A&S; Andrew Gerber, Assistant Professor of Psychiatry, CUMC, and Director of the New York State Psychiatric Institute's Magnetic Resonance Imaging Program; and Kyle Mandli, Assistant Professor of Applied Physics & Applied Mathematics, SEAS, the Subcommittee includes representation from A&S, SEAS, and CUMC. In addition, Bruno Scap, Assistant Director of Internet Technology for the Data Science Institute and the Department of Biomedical Informatics, CUMC, has joined the Subcommittee as an invited observer. Bose reviewed the Subcommittee's earlier Fall 2014 recommendations:

- **(Done)** Continue the Subcommittee into Spring with charge to investigate resource sharing;
- **(Done)** Expand membership to include A&S, SEAS, and CUMC-IT leadership;
- **(In Process)** Support Cloud Subcommittee recommendation for a university enterprise agreement for Amazon and other services;
- **(In Beginning Stages)** Begin a coordinated effort to establish a Columbia "Research Computing/IT Staff" community;
- **(To Be Decided)** Dedicate a staff person to identify and track university options for external resources;
- **(To Be Decided)** Survey new Data Science Institute faculty for research computing power consumption to measure demand for resources.

The Subcommittee added two new recommendations:

- Create a Columbia "Research Computing" portal, wherein researchers can find information such as what software licenses exist, courses that can be taken, and so forth;
- Identify solutions for affordable and remotely backed-up research storage options and affordable space for hosting research computing equipment on campus. Chair Johnston indicated that the issue of data storage was a guiding force behind the creation of the new [Research Data Advisory Committee](#) to be discussed later in the agenda.

Bose added that the Intercampus Subcommittee also emphasized the importance of the ability of researchers to use available shared computational resources regardless of the campus on which the

researcher or resource resides. In addition, the Subcommittee discussed identifying existing resources that are not yet shared, but could be. Johnston thanked Bose for his presentation and his work, and asked that the Subcommittee continue in the FY16 year. [Presentation attached as Appendix I.](#)

Yeti Operating Committee

Professor Greg Bryan gave the Yeti Operating Committee report. He noted that the major accomplishment was the successful installation of Yeti II, which expanded the system by 66% in February 2015. Usage has increased substantially, and there are now 164 users from 24 groups who have used Yeti for 1.2M jobs since January 2015. Professor Jerry Ostriker asked if this was a significant increase; Bryan responded that it was with the number of groups that bought into the cluster approximately doubling. Bryan pointed out that although the number of cores increased by 66% from Yeti I to Yeti II, the number of cores purchased by faculty doubled as the first round included cores purchased through a grant secured to help build the facility.

Bryan asked that the community continue to inform him or RCS of any publications that resulted from Yeti as it helps make the case for its importance.

Bryan then said that the Yeti OC had a very productive meeting recently to consider various options for management policies. Twenty people attended, including all six Executive Committee members and six CUIT staff. Bryan reminded Committee members that the format of the Yeti Operating Committee is that anyone interested may join for the first 50 minutes for an open discussion, followed by a brief Executive Committee meeting. Three changes were made: (a) added parallel queue for MPI jobs; (b) software updates; and (c) operational changes. [Slides](#) and [minutes](#) from the Spring 2015 Yeti Operating Committee meeting are posted on the SRCPAC web page. [Presentation attached as Appendix II.](#)

Education Subcommittee

Chair Johnston then spoke on behalf of Professor Bob Mawhinney, who was teaching during the scheduled meeting time, and was thus unable to attend. She reminded the Committee that at its Fall 2014 meeting, members approved setting aside up to 5% of capacity for HPC teaching. Professor Mawhinney anticipates using this newly-established resource for his Fall course. The Committee asked whether any other class was currently using the resource, and how professors/instructors were informed of the resource's availability. Victoria Hamilton responded that it was a challenge to communicate broadly, but that a subgroup of faculty involved in teaching data science had met twice in the Spring 2015 semester, and there was substantial evidence of an increasing number of data science classes across the university, many tailored to specific schools and disciplines.

Johnston also mentioned the workshops for Introduction to Linux and Introduction to HPC, which had been held throughout the academic year; Johnston asked Rob Lane if these workshops would be continued; to which Lane assured the committee that they would.

Cloud Pilot Summary

With the conclusion of the Subcommittee reports, Chair Johnston asked Rob Lane to update the committee about the Cloud Pilot recommended at the Fall 2014 SRCPAC meeting. Lane reported that in response to an email to all SRCPAC attendees, six Cloud Pilot projects had been suggested. RCS was preparing to pilot one of them in the near future – a proposal from the Data Sciences Institute that focused on deploying Natural Language Processing to analyze news streams. Professor Chris Marianetti asked why there would be just one pilot, as there were many ways to use the Cloud. Lane responded that RCS wanted to start with one to focus on ascertaining whether there were any mechanical or administrative issues to resolve. Professor Jerry Ostriker asked about procurement policies. Chair Johnston said that would be addressed as part of a later agenda item.

Research Data Advisory Committee

Amy Nurnberger, Research Data Manager in the Library's Center for Digital Research & Scholarship said she had formed a committee to focus on the increasing need for Data storage and curation, particularly in light of new demands by funders and publishers. The inaugural meeting was planned for the following day. The committee is starting with approximately 10 people, but would expand. An attendee asked if the committee was interested in storage and curation only after work was completed or also during the research process? Nurnberger said that was one of the issues before the inaugural committee. Strong interest in this important issue was voiced by a number of attendees and encouragement to review the entire [Data Life Cycle](#).

Potential for HPC Capacity/Amazon Enterprise Agreement

Alan Crosswell, Associate Vice President & Chief Technological Officer within CUIT, reported that progress was being made on the Amazon Enterprise Agreement to ensure that the offering conformed to the needs of the community for issues such as security, privacy, and accessibility. He also indicated that CUIT was exploring a variety of options to expand capacity for HPC should it exceed the current 100 kW capacity. Among the options were expanding cooling, linking various Columbia Data Centers together, joining the Massachusetts Consortium, etc. Gaspare LoDuca, Columbia's new Chief Information Officer and Vice President within CUIT, reinforced Crosswell's remarks, indicating that CUIT was both exploring short-term options and planning to develop a longer-term strategic plan. The remarks were welcomed by the SRCPAC attendees.

Faculty Recruitment Guidelines & Recommendations

Chair Johnston then circulated a proposed set of guidelines and called for discussion about how to balance allocating capacity among (a) the normal annual purchase cycle, typically in the Fall; (b) recruits, typically in the Spring; and (c) proposals that incorporate substantial purchases, which happen throughout the year. She spoke from personal experience as the Chair of Astronomy, a department that is recruiting two candidates this Spring. She said that the existence of the SRCF and shared HPC had been critical to each, but that the nature of recruitment often requires acting very quickly to respond to counter-offers. Faculty agreed as to the time-critical nature of recruitment offers. Professor Marianetti also gave by way of example the fact that he had recently submitted a grant that incorporated in its budget \$500k for computational resources. After discussion, Johnston thanked attendees for their suggestions and said she would be presenting the suggested policy to the Research Computing Executive Committee (RCEC) in May. [Guidelines attached as Appendix III.](#)

Next Purchase Cycle

Chair Johnston called on Victoria Hamilton, Director of Research Initiatives within the Office of the EVP for Research, to outline the next purchase cycle. Hamilton indicated that while originally it had been thought that a biannual purchase cycle made sense (as that is what other schools often do), last year's surge in interest, just a year after the start of Yeti I, suggested that ascertaining interest should be done annually. It also appeared a September deadline best suited recruits and departments alike.

Reflecting discussion during the Yeti II purchase, Hamilton said a request for an email of intent would be sent to Chairs and SRCPAC members in May to determine whether there was sufficient volume to justify forming a faculty committee to advise on the next RFP. Professors Ostriker, Marianetti, and Johnston felt it was highly likely that demand would be sufficient, and that the committee should be formed without the need for the email. Alan Crosswell agreed as the previous RFP had taken place some years previously for the first round of Yeti. Chair Johnston suggested that any large purchasers should

automatically be invited to serve on the Advisory Committee and also called for volunteers. Professor Marianetti agreed to serve.

Upcoming RCEC Meeting & Closing Remarks

Chair Johnston asked if there were any further issues to discuss, thanked the attendees for their participation, and reminded them that she would be reporting the accomplishments and recommendations of SRCPAC to the Research Computing Executive Committee (RCEC) in May. The meeting was adjourned at 12:55 p.m.

Appendices

[Appendix I](#) – Intercampus Subcommittee PowerPoint Presentation (Raj Bose)

[Appendix II](#) – Yeti Operating Committee PowerPoint Presentation (Greg Bryan)

[Appendix III](#) – Faculty Recruitment Guidelines & Recommendations (Kathryn Johnston)