COLUMBIA | RESEARCH
Shared Research Computing
Policy Advisory Committee

SRCPAC Minutes - Spring 2023
Date: Tuesday, February 6
Time: 2:30-4:00
Meeting called by: Alex Urban, Chair of SRCPAC

MEETING MINUTES

Welcome & Introductions – Alex Urban, Chair of SRCPAC

Alex Urban, Chair of SRCPAC, called the meeting to order and proceeded to outline the agenda. He welcomed Hod Lipson, Co-Chair of the Research Computing Faculty Committee (RCFC), to provide updates.

RCFC Report – Hod Lipson, Co-Chair of RCFC

Professor Lipson introduced himself and proceeded to remind the committee of the RCFC charge. RCFC was formed to recommend a strategic plan for the University's future computational and data infrastructure for research, with the goal of submitting a report by the end of the year (May 31, 2024).

Professor Lipson explains that the impact of AI and high performing computing science will affect all angles of the university. The presumption of this only affecting mostly science and engineering is no longer true – every discipline field is being transformed by AI.

Upon forming the committee, the chairs sent out a survey to researchers asking "what research computing infrastructure would researchers require in the next decade to flourish, compete, and lead in your field?"

Going into detail on the evolution of shared computing resources, Professor Lipson details the needs going forward: advanced users need cost-effective HPC resources, novices need training, and many existing baseline infrastructure needs need to be resolved. Outlined in the format of Maslow's hierarchy of needs, Research computing hierarchy of needs is outlined as such (bottom to top):

- Infrastructure needs – reliable connectivity, workstations, power
- Basic services – email, security, storage, backup, websites, admin
- Skilling needs – research computing training, courses, resources
- Guided exploration – software tools, sandboxes, datasets, models
- Self-driven research – raw HPC compute power and large storage

CUIT is focused on bottom layers of infrastructure and basic services. Research computing is focused on the top layers of skilling, guided exploration, and self-driven research.

He proposed the idea of a "Discovery Accelerator," focusing on establishing a lean research computing center to provide access to high-performance computing and education, with a specific fundraising target. The support for this initiative would need to come directly from departments.

Professor Urban then introduced Jeannette Wing, Executive Vice President for Research, to discuss the recent Empire AI Announcement.

COLUMBIA | RESEARCH
Shared Research Computing
Policy Advisory Committee

Empire AI Announcement – Jeannette Wing, Executive Vice President for Research

Professor Wing updated the committee about the Empire AI initiative proposed by Governor Hochul's office, aimed at leveraging AI technologies for governance, public services, and economic development in New York State.

Professor Wing hands the meeting back over to Professor Urban.

High-Performance Computing Update – Alex Urban, Chair of SRCPAC

Professor Urban provided an overview of SRCPAC's origins and its focus on policy issues related to shared research computing. He reviewed the current HPC footprint, noting constraints in cooling, power, and staff resources.

Professor Urban reviewed the current HPC footprint and status of current clusters as further described in the **attached slides.**

Current Clusters

- Terremoto Phase 2
- Ginsburg Phase 1, 2, 3
- Manitou – GPU Cluster
- Insomnia (most recent addition to replace computer cluster)
- Free Tier (accessible in every cluster)

In the past year, there have been 1,129 users on Terremoto and 1,070 on Ginsburg. Users span all over campus.

He discussed the phased retirement of existing clusters and the introduction of Insomnia, a new cluster following a different purchasing model. Communication regarding this new model will be forthcoming from CUIT.

Professor Urban then passed on to his colleague Axinia Radeva, Manager of CUIT Research Services, for updates on computing services.

Research Computing Services Update - Axinia Radeva, Manager of CUIT Research Services

Axinia Radeva presented updates on research computing services, including new offerings such as Overleaf Professional. She highlighted enhancements to existing services like Secure Data Enclave and efforts to expand access to resources like Globus.

Current Research Computing Services

- Embedded research computing support
- Secure data enclave (SDE)
- Electronic research notebooks with LabArchives
- Globus

- Cloud Research Computing Consulting
- Access National HPC Campus Contact (Former XSEDE)
- SnapGene
- Overleaf professional (NEW)

More information can be found in the **attached slides**.

CUIT has updated their webpages for your CUIT research resources and trainings:

Link to page: https://www.cuit.columbia.edu/research
Video library of training recordings: https://www.cuit.columbia.edu/rcs-videos

To subscribe to the EVPR and Columbia Libraries Researcher Resource newsletter, follow this link.

Overleaf Professional launched last August, providing a centralized license for all Columbia users. Since acquiring, there has been a 175% increase in users.

Secure Data Enclave added a Linux environment in September 2023 and had a hardware upgrade last year.

CUIT is currently working with CUIMC to get Globus access for all Columbia users.

XSEDE now known as ACCESS is an NSF-funded, nationwide collection of supercomputing systems available to researchers through merit-based allocations. In September 2023, ACCESS approved a supplement of 750,000 ACCESS credits to our Columbia Discover Allocation, which is used for small scale testing and benchmarking for researchers.

GraphPad prism group subscription license is available to Columbia users. It launched last week with 200+ users.

Concluding with RCS contact information, Ms. Radeva turned it over to Marc Spiegelman, Chair of Foundations for Research Computing, to update SRCPAC.

Foundations for Research Computing Update - Marc Spiegelman, Chair of the FRC Advisory Committee, Anne Cong-Huyen, Jonathan Cain

Professor Spiegelman discussed Foundations for Research Computing, starting with the mission statement.

Foundations for Research Computing provides an informal introduction for Columbia University graduate students and postdoctoral scholars to the fundamental skills for harnessing computation: core languages and libraries, software development tools, best practices, and computational problem-solving.

Professor Spiegelman laid out Foundations' Primary activities:

- Novice trainings
- Data Club
- Workshops

Foundations has worked to apply to different levels through experimenting with various trainings and workshops.

In the past year, there has been two new hires – Anne Cong-Huyen (Director of Digital Scholarship) and Dan Woulfin (Computational Research Instruction Librarian). Foundations now lives in the Libraries department.

Professor Spiegelman emphasizes again that the issues remain the same: the high demand for teaching at novice level and inability to scale.

Looking at the past 5 years of bootcamps, we can see where we've done well and where we need help.

In January 2023, there were 188 applicants with 28 accepted attendees. By January 2024, demand returned at 508 applicants and 58 accepted attendees. Most applicants over the years were masters students, spanning various schools but mostly from GSAS and SEAS.

Professor Spiegelman addresses the main question: how to scale to meet demand? We've seen that the current model is effective but labor intensive requiring 2 instructors and 5 helpers per 30 people. All staff are volunteers.

Summarizing more issues moving forward,

- How do we maintain contact with faculty and their changing needs for their students?
- Are there alternatives to the labor-intensive model and curriculum we currently use?
- How can we better serve the increasing numbers of intermediate learners?

To evolve, unmet needs need to be addressed to develop a more sustainable, expanded Foundations program.

Professor Spiegelman passed this on to Professor Urban.

Discussion of Future Directions – Alex Urban, Chair of SRCPAC

Professor Urban initiated a discussion on the current status of SRCPAC, highlighting both its successes and areas for improvement.

He noted that the access policies for Shared HPC have proven effective, and concerns regarding buy-in and high costs have been mitigated through the implementation of new purchasing models and options. Additionally, various research computing services have become more accessible to Columbia users.

Identifying areas for enhancement, Professor Urban emphasized the need to improve communication and marketing strategies, enhance entry-level training programs, address computing needs comprehensively, maintain free-tier options, and bolster faculty involvement in SRCPAC initiatives.

Regarding faculty engagement, Professor Urban encouraged committee members to consider nominating themselves or a colleague for the position of SRCPAC co-chair or to participate in a temporary ad-hoc subcommittee focused on entry-level training.

Other Business & Closing Remarks – Alex Urban, Chair of SRCPAC

Professor Urban then invited further comments and questions from the attendees.

Professor Ying Wei raised a query regarding research computing for medical data. Ms. Radeva addressed the concern, stating that CUIT acknowledges the necessity for HPC resources for the medical campus and is collaborating with CUIMC's IT department to find a suitable solution, considering the constraints associated with handling medical data. The current focus lies on improving file transfer capabilities with Globus.

Professor Jeannette Wing inquired about addressing the RCFC report in the absence of proper funding. She questioned whether the issues could be managed with existing resources. CUIT is expected to play a central role in management, albeit not in domain-specific areas. Local peers would offer assistance, leveraging the established network.

Another committee member sought clarification on the new buy-in process with Insomnia and expressed concerns about potential quality issues if other users access the nodes. Professor Urban assured that quality concerns should not arise, and the administration of purchases would be streamlined.

Another member inquired about Columbia's power expansion plan to support future growth. While the university currently lacks the capacity for high-power infrastructure, potential solutions such as carbon neutrality and leveraging local infrastructure are being explored. However, there is no definitive solution at present, though future developments may yield solutions.

With no further questions from the Committee, Professor Urban expressed gratitude for everyone's contributions and participation, and concluded the meeting.

Slides provided on https://research.columbia.edu/content/srcpac

See full list of attendees below.

Attendees:

Maneesha Agarwal	*Owen Keith*	*Dustin Rubenstein*
Timothy Berkelbach	*Kriste Krtovski*	*Lorenzo Sironi*
Jonathan Cain	*Rob Lane*	*Marc Spiegelman*
Rob Cartolano	*Hod Lipson*	*Anthony Vanky*
Kitty Chan	*Gaspare Loduca*	*Eric Vlach*
Seth Cluett	*Mahdad Parsi*	*Ying Wei*
Roisin Commaine	*Darcy Peterka*	*Tian Zheng*
Alan Croswell	*Axinia Radeva*	*+ 8 more*
Michael Faucher	*Murdach Reilly*	
Jeff Goldsmith	*David Romoff*	

(35 attendees total)