

**SRCPAC Minutes - Spring 2023**

**Date:** Monday, April 24, 2023

**Time:** 10:35-12:00

**Meeting called by:** Chris Marianetti, Chair of SRCPAC

**MEETING MINUTES**

**Welcome & Introductions - Chris Marianetti, Chair of SRCPAC**

SRCPAC Chair, Professor Chris Marianetti, called the meeting to order and introduced the new co chair, Professor Alexander Urban, who will be taking over as SRCPAC chair July 1 of this year. He notes that this is our first meeting back in-person in three years. Following the usual protocol, he directed participants around the room to introduce themselves. Professor Marianetti then proceeded to outline the agenda and welcomes EVP for Research, Jeannette Wing, to provide updates on the development of a university-wide long-term plan for computing and data storage.

**Long Term Strategic Thinking about University Needs for Computing and Storage - Jeannette Wing, EVP for Research and Professor of Computer Science**

EVP Wing introduced herself and explained how Professor Marianetti's report to the RCEC last year sparked the initiative to think about a long- term strategy to provide the university research community the compute and storage it will need in the future. She thanked Professor Marianetti for his impact as chair especially in general and specifically for stimulating this effort to think about the next steps for Columbia: where is research computing infrastructure going?

EVP Wing recruited Gaspare LoDuca, CIO/VP for Information Technology as her co-chair. Since last September, Professor Wing and Mr. LoDuca have met with the research deans and heavy users in schools including the CUIMC schools, SEAS, SIPA, Law, Business, Architecture, Climate School, and DSI, along with some chairs for Arts and Sciences. There was unanimous buy-in for developing a university strategy.

CIO LoDuca and EVP Wing asked Maneesha Aggarwal, AVP for Academic, Emerging Technologies & Research Services and CUIT, and Sophie Thuault-Restituuto, EVPR Chief of Staff and Executive Director for Special Projects to staff this effort and meet with faculty groups for their input on their anticipated future computing needs are. Though collected information is already outdated due to ChatGPT and the advances of last few months, The team compiled the input and data (although EVP Wing pointed out that the emergence of ChatGPT has already made those estimates obsolete.) In January, a Faculty Committee formed, and will meet on a regular basis with the intention of delivering a recommendation by end of fiscal year 2024.

The Faculty Committee consists of:

- **Timothy Berkelbach**, Associate Professor of Chemistry; Faculty of Arts and Sciences



Shared Research Computing  
Policy Advisory Committee

- **Roisin Commane**, Assistant Professor of Earth and Environmental Sciences, Atmospheric Composition Group, Lamont Doherty Earth Observatory; Faculty of Arts and Sciences
- **Wojciech Kopczuk**, Professor of Economics and of International and Public Affairs; Faculty of Arts and Sciences and School of International and Public Affairs
- **Hod Lipson**, co-chair, James and Sally Scapa Professor of Innovation in the Department of Mechanical Engineering; Co-Director, Maker-Space Facility; School of Engineering and Applied Sciences
- **Ciamac Moallemi**, William von Mueffling Professor of Business; Columbia Business School
- **Darcy Peterka**, co-chair, Senior Research Scientist; Scientific Director of Cellular Imaging; Mortimer B. Zuckerman Mind Brain Behavior Institute
- **Muredach Reilly** (interim), Florence and Herbert Irving Endowed Professor of Medicine; Director, Irving Institute for Clinical and Translational Research; Associate Dean for Clinical and Translational Research; Columbia University Irving Medical Center

Ex-officio members are:

- **Robert Cartolano**, Associate Vice President for Technology and Preservation, Columbia University Libraries
- **Gaspere LoDuca**, Chief Information Officer and Vice President for Information Technology, Columbia University Information and Technology
- **Alexander Urban**, Assistant Professor of Chemical Engineering, School of Engineering and Applied Sciences and Chair of the Columbia Shared Research Computing Policy Advisory Committee (SRCPAC)
- **Jeannette M Wing**, Executive Vice President for Research; Professor of Computer Science, Office of Research

Staffing the committee will be:

- **Maneesha Aggarwal**, Assistant Vice President, Academic, Emerging Technologies & Research Services, Columbia University Information and Technology
- **Sophie Thuault-Restituto**, Chief of Staff and Executive Director for Special Projects, Office of Research

Professor Tian Zhang expressed her concerns regarding the scope of the charge, which she argued would benefit by incorporating the educational and social missions of a university which also use compute and data storage infrastructure. EVP Wing responded that this concern would be conveyed to the Faculty Committee in the next meeting.

Professor Wing turned the meeting back over to Professor Marianetti, who in turn introduced the Chair of the HPC Operating Committee, Kyle Mandli.

**High-Performance Computing Update - Kyle Mandli, Chair of the HPC Operating Committee, John Villa, Manager of High-Performance Computing, CUIT**

Professor Mandli had no updates regarding the HPC Operating Committee and asked John Villa, the new Manager of High-Performance Computing, to report on the status of the HPC clusters and the need for an RFP committee to evaluate alternatives for the next cluster. Mr. Villa replaced Cesar Arias a few months ago.

## Status of Current Clusters

Mr. Villa reviewed the status of current clusters as further described in the **attached slides**.

- Habanero has been moved out of the high-density racks within the Data Center
- Habanero's Phase 2 will retire in May and phase 2 users will be moved to the free tier
- Terremoto Phase 1 will retire in December 2023 and Phase 2 in December 2024.
- Ginsburg Expansion 2 went live in December 2022. The cluster will total 286 nodes, 9152 cores and 39 GPU hardware accelerated systems.

## GPU Cluster – Manitou

Manitou was delivered in late February 2023 and is currently live. Expansion 2 is experiencing supply chain issues.

## RFQ Committee

The RFQ Committee convened in early March. Those that attended included Kyle Mandli, Julia Hirschberg, Bob Mawhinney, and Rob Lane. The committee agreed on specifications that were sent to 15 vendors and narrowed down to 5 finalists. The open order period should start May 15 – June 15. In the next phase of HPC, the RFQ committee hopes to focus on centralized storage, provisioning, and service.

Mr. Villa then passed the microphone 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 started with the goals for Research Computing Services:

- Expand RCS portfolio
- Support research activities across Columbia University

Specific research computing services include:

- 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

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

Overleaf Professional will be available in July. CUIT and the Libraries are partnering to provide a license for all Columbia users which should save users money. Transition from personal to university accounts should be seamless.

The Embedded Research Computing support has added two new positions.

The Secure Data Enclave is a secure, centralized service for faculty and researchers that works with sensitive research data. The current on-premises environment will be upgraded due to increasing demand SDE in-the-cloud does not have enough use cases.

The number of users of SnapGene doubled since Spring 2022. A webinar was held in February 2023 offering Columbia users insights directly from a field scientist at SnapGene.

Ms. Hescok then asked Maneesha Aggarwal, AVP of Academic, Emerging Technologies & Research Services for Columbia University Information and Technology to speak to new projects.

### **Data Catalog Project and the Proof Of Concept with school of Nursing/Newsletter for Researchers - Maneesha Aggarwal, CUIT AVP, Academic, Emerging Technologies & Research Services**

Ms. Aggarwal, introduced herself and presented the new Data Catalog Project headed by CUIT, the Libraries and EVPR. The project hopes to create a data repository that automates data from a Qualtrics form to create a widely available source. The project is being tested with the School of Nursing. If successful, they hope to expand it to other schools.

Ms. Aggarwal is also working on restoring the quarterly Newsletter for Researchers. Partnering again with the Libraries and EVPR, they hope to publish a quarterly newsletter with sections from each department with updates regarding research resources and updates.

Ms. Aggarwal turned over the meeting 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, Jeremiah Trinidad-Christensen**

Professor Spiegelman discussed Foundations for Research Computing, starting with the mission statement which focuses on graduate students and postdocs.

Professor Spiegelman laid out Foundations' Primary activities:

- Novice trainings
- Data Club
- Workshops

Since 2018, there have been 12 bootcamps with 2-3 per year. The last bootcamp was on January 12-13, 2023 and returned in-person after nearly three years. As in previous years, Demand continues to far exceed supply as the bootcamp could only host 31 researchers from the 300 applications distributed over schools from across the university

Professor Spiegelman then introduced Data Club . Sessions have been occurring every two weeks; recordings are available on Youtube. This successful model taps into knowledge and talent within the University not normally utilized, providing a venue for students to help students.

Professor Spiegelman then addressed two key questions for Foundations's future: high demand and inability to scale. The demand has always exceeded supply. Moreover, SRCPAC only targets postdocs/grad students, but other groups also want this training. Since Foundations now resides within the Libraries, whose mandate is to support the entire University, should Foundations scale to meet demand? If so, how? Today the program depends on volunteers for the workshops. Rob Cartolano commented on the use of developers as volunteers for program. Though it is a great opportunity for instructors to meet students, there is a limit to how much they can volunteer. CTO Alan Crosswell commented that his staff likes volunteering at the workshop as it connects them with the students and research at the heart of the university's mission, but agreed that a scaling problem exists. Funding, support, and full-time personnel/oversight are needed to take these programs further.

The Libraries are currently searching for two roles: Computational Research Instruction Librarian to replace the Foundations Manager that left nearly two years ago, and a Director for Digital Scholarship to whom the Librarian would report

After further discussion, Professor Spiegelman passed this on to Professor Marianetti.

### **Other Business & Closing Remarks - Chris Marianetti, Chair of SRCPAC**

Professor Marianetti opened the floor to further comments and questions.

Ms. Hescock thanked Professor Marianetti for his unparalleled service as SRCPAC chair.

With no further questions from the Committee, Professor Marianetti thanked everyone for their input and participation and closed the meeting.

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

*See full list of attendees below.*

#### **Attendees:**

<i>Mohammed AlQuraishi</i>	<i>Axinia Radeva</i>
<i>Rob Cartolano</i>	<i>Marc Spiegelman</i>
<i>Alan Crosswell</i>	<i>Sophie Thuault-Restituto</i>
<i>Kitty Kay Chan</i>	<i>Joyee To</i>
<i>Jessica Eaton</i>	<i>Jeremiah Trinidad-Christensen</i>
<i>Victoria Hamilton</i>	<i>Alexander Urban</i>
<i>Halayn Hescock</i>	<i>John Villa (Zoom)</i>
<i>Lokke Highstein (Zoom)</i>	<i>Tian Zheng</i>
<i>Rob Lane</i>	
<i>Kyle Mandli</i>	
<i>Chris Marianetti</i>	
<i>Razvan Popescu</i>	