

SRCPAC Minutes - Fall 2022

Date: Friday, December 2, 2022

Time: 2:00- 3:20

Meeting called by: Chris Marianetti, Chair of SRCPAC

MEETING MINUTES

Welcome & Introductions - Chris Marianetti, Chair of SRCPAC

SRCPAC Chair, Chris Marianetti, called the meeting to order and outlined the agenda. He regrets the need to hold this meeting virtually, and welcomes EVP for Research Jeannette Wing and Chief Information Officer Gaspare LoDuca to speak about their efforts to develop a university-wide long-term plan for compute and data storage after

Halayn Hescoock, Senior Director of CUIT Research Services gives an overview on Data Center Capacity.

Shared Computer Capacity - Halayn Hescoock, Sr Director of CUIT Research Services

Halayn reviewed the capacity of the Shared HPC Data Center, which is a function of physical space, electricity, cooling, as well as advancing capabilities of new equipment. Assuming we continue at the same pace of growth, we could fill two racks per year and accommodate growth as old clusters roll off.

Halayn then passed on to Jeannette Wing, EVP for Research, and Gaspare LoDuca, CIO/VP for Information Technology of CUIT for their thoughts on long term plans for the University's need for computing and storage.

Long Term Strategic Thinking about University Needs for Computing and Storage - Jeannette Wing, EVP for Research, Gaspare LoDuca, CIO/VP for Information Technology, CUIT

Professor Wing introduced herself and explained that Chair Marianetti's annual report to the RCEC sparked the need to think now about the future of Columbia's research computing and data storage as SRCPAC will reach a limit at some point. This provides a great opportunity for the university to think about long term computing and data needs for entire university.

This idea, presented to the Council of Deans, generated a huge amount of enthusiasm to explore a big vision for research computing. Since September, Gaspare and Jeannette have met with the research deans and heavy users in schools including the medical school, SEAS, SIPA, Law, Business, Architecture, and Climate School.

Next Sophie Thuault-Restituuto, EVPR Chief of Staff and Executive Director for Special Projects, and Maneesha Aggarwal, CUIT AVP of Academic, Emerging Technologies & Research Services, will meet with faculty to hear what their needs are and create a faculty committee to develop a set of recommendations to central administration. It seems likely that a long-term plan/solution will combine on-site and cloud storage. Moreover, support is not just about storage and computing but about the staff.

 COLUMBIA | RESEARCH
Shared Research Computing
Policy Advisory Committee

Mr. LoDuca added that these efforts aim to make sure every researcher can compete on grants that require computing. Researchers should not lose grants due to lack of resources. He emphasized the critical role the faculty will play on this committee to help voice their needs and how necessary growth is.

Professor Marianetti commented that SRCPAC has called out the need for this kind of thinking for some time, and it's great that Professor Wing and Mr. LoDuca are creating a vision that indicates we need investment.

Professor Wing responds that they raised the need for investment with the deans but the deans need to hear from the faculty and to be involved.

Professor Wing turned the meeting back over to Ms. Hescock, who in turn introduced the Chair of the HPC Operating Committee, Kyle Mandli, to summarize the committee's activities.

High-Performance Computing Update - Kyle Mandli, Chair of the HPC Operating Committee, Cesar Arias, Manager of High-Performance Computing, CUIT

Professor Mandli spoke briefly about the Operating Committee and indicated the next meeting will take place in January for all HPC users. Professor Mandli then asked Cesar Arias, 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.

Status of Current Clusters

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

- Habanero's Phase 1 was retired at the end of May 2022. As its hardware is no longer under warranty, nodes will be disposed of as they die. Phase 2 will retire in April of 2023.
- Terremoto Phase 1 will retire in December 2023 and Phase 2 in December 2024.
- Ginsburg Expansion went live in March 2022, expanding to 99 nodes with an additional 3168 cores. Ginsburg Expansion 2 is anticipated to be available by end of year 2022. The cluster will total 286 nodes, 9152 cores and 39 GPU hardware accelerated systems. Ginsburg use has been trending down the past few months as the slowness issue affected job completion numbers and run time.

Shared high performance computing in the past ten years has grown exponentially with more than 18 million jobs run, 363 million core hours of compute provided, 350 peer-reviewed publications, and users from over 70 groups and departments.

GPU Cluster – Manitou

Supply chain issues have pushed the delivery of the new cluster 15 GPU node Manitou to January 2023. .

RFP committee

COLUMBIA | RESEARCH

Shared Research Computing Policy Advisory Committee

Mr. Arias then asks for volunteers to be a part of the RFP Committee that will focus on new storage solutions/vendors for the next purchase. There will be 4 meetings over 2-3 months starting Feb/Mar 2023. Ms. Hescoek will reach out with an email requesting for participation. Mr. Arias opened up for questions.

Rob Lane noted that when vendors delayed some past orders, the vendor offered to over deliver. He asked if this could be raised with vendors. Mr. Arias offered to further discuss with Mr. Lane.

Mr. Arias 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 touched on the following:

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
- LabArchives
- Scientific software
- SnapGene

The Embedded Research Computing support is a growing service and is getting more attention. CUIT hires and trains individuals who on-site research computing support at Columbia research departments or centers. New staff has been added since the beginning of the year.

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

The LabArchives Forum this year occurred on both campuses and allowed users to meet experts in LabArchives in-person. CUIT and the Libraries jointly fund this service so is available at no cost to users. With over 30+ integrations, LabArchives is mostly used at CUIMC but several administrative units are also leveraging the tool.

Axinia then summarized the results of an RCS embedded Engineers scientific survey conducted in May which include:

- The number of SnapGene users increased by 15%
- 176 group licenses with a total of 891 activations between licenses and a total of 434 personal subscriptions show the demand for a GraphPad Prism bulk license account

CUIT and the Libraries are working on an enterprise license for Overleaf as 9017 confirmed Overleaf users registered. Statistics show Columbia's users often come from Computer Science, Physics,

 COLUMBIA | RESEARCH
Shared Research Computing
Policy Advisory Committee

Economics, Electrical Engineering and Mathematics, but only a small sample of users completed this field.

Professor Marianetti asked why not use GitHub and asked if there is a better way to systemize the process for getting an institutional license. Ms. Aggarwal responded that her team would look more into having a more transparent place for suggesting software.

Ms. Hescock then asked Marc Spiegelman, Chair of Foundations for Research Computing, to update SRCPAC.

Foundations for Research Computing Update - Marc Spiegelman, Chair of the FRC Advisory Committee, Jonathan Cain, Associate University Librarian for Research and Learning

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

Professor Spiegelman emphasized that Foundations and the Libraries are still looking for a Central Coordinator after more than a year.

The next bootcamp is scheduled for in-person in January 2023. Demand continues to far exceed supply for Foundation's two-day workshops. The Fall 2022 bootcamp held remotely received 423 applications distributed over all schools and degree programs but focused on researchers.

Professor Spiegelman then introduced Data Club – a pathway from bootcamp to expand from the bootcamp. It offers a coworking space to share and discuss issues. For example, the sessions in fall focused on Python.

Foundations Partner workshops were offered remote and in-person. With the same basic offerings, there was the similar issue of high demand. Admirable work was being done to keep it moving but exacerbated by lack of the program coordinator.

Professor Spiegelman stated that Foundations and Libraries converted the coordinator role to a computational research Librarian position in an attempt to increase strength of the recruitment pools.

Professor Spiegelman explained that the problem is the demand has always been higher than what Foundations can provide. Data from bootcamps shows that 40% of the applicants are post-docs and doctorates. In terms of future goals, Foundations hopes to recruit the Program Librarian, continue/expand current programs, develop of a more robust assessment plan to address scaling issue, and seek input/new partners in guiding the future of Foundations.

Professor Spiegelman then turned to the audience for questions. Rob Cartolano commented on using developers as volunteers for program as dedicated volunteers lose edge for developing. Though it is a great opportunity for instructors to meet students, there is a limit to how much they can volunteer.

Professor Marianetti responds that the bigger picture is to hire a new coordinator that will help figure out solution for better support. All ties into same issue – the program can't scale on just volunteer efforts.

 COLUMBIA | RESEARCHShared Research Computing
Policy Advisory Committee

Professor Spiegelman then asked Jonathan Cain, Associate University Librarian for Research and Learning, and Jeremiah Trinidad-Christensen, Head of Research Data Services, for additional comments. Mr. Trinidad-Christensen responds that the answer isn't more of the same thing and that they are trying to find other ideas to support the folks that are interested in attending bootcamps.

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

Professor Marianetti opened the floor to further comments and questions.

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

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

See full list of attendees below.

Attendees:

Joyee To	Tian Zheng
Jonathan Cain	Tameek Henderson
Marc Spiegelman	Jessica Eaton
Axinia Radeva	Kitty Kay Chan
Cesar Arias	George Garrett
Halayn Hescoock	Winnie OKelley
Kyle Mandli	Mahdad Parsi
Victoria Hamilton	Zsuzsa Marka
Chris Marianetti	John Hunt
Jeremiah Trinidad-Christensen	Rob Lane
Razvan Popescu	Al Tucker
Abadali Sheikh	Alexander Urban
Alexander Bergier	Christian Moser
Rob Cartolano	Gaspere LoDuca
Lokke Highstein	Harmen Bussemaker
Alan Crosswell	Jacqueline Austermann
John Villa	Maxfield Shortte
Julia Hirschberg	Mike Faucher
Michael Raghunandan	Paul Blaer
Mohammed AlQuraishi	Rob Lane
Razvan Popescu	Robert Kudyba
Sophie Thuault-Restituto	Roger Creel
Timothy Berkelbach	