Engineering at Illinois IT Governance Research Working Group

Final Recommendations and Report on Activities during the 2024-2025 Academic Year

Members:

Andre Schleife (schleife@illinois.edu, MatSE) - Chair
Nick Colaneri (nfc@illinois.edu, GCOE office of Research)
Mark Neubauer (msn@illinois.edu, Physics)
Jianqi Xi (jianqixi@illinois.edu, NPRE)
Jinhui Yan (yjh@illinois.edu, CEE)
Mark Hart (mshart@illinois.edu, Engineering IT – Ex Officio)

Executive Summary

The working group makes the following recommendations:

- GCOE Leadership should pursue the development of NPCF as a shared datacenter based on a proposal delivered to Dean Bashir in March, 2025.
- Continue exploration - through departmental stakeholder engagement - of the value a dedicated software specialist would add to the college with an expectation of proposing the creation of this position in the next year.
- Continue efforts to emphasize key distinctions between data storage locations (both cloud and conventional), and continue to advise staff, faculty, and students on locations and strategies to both store and preserve their data long-term.
- Continue to monitor the development of AI tools, services, and policies to look for opportunities and solutions that make sense for investment on behalf of GCOE faculty.
- Continue ongoing efforts to increase the awareness of resources for GCOE researchers and evaluate the effectiveness of those efforts.
- Approve a $20,000 investment in Engineering IT's supplemental/test GPU system to ensure it meets the evolving needs of GCOE researchers.
- Consider a $275,000 investment in GCOE's Campus Cluster Investment. This would provide GCOE a dedicated node with advanced H200 GPUs to meet increasing research demands in the AI/ML space, while affording the college the flexibility of redirecting existing cluster investments towards instructional use cases.
- Continue working with Illinois Computes to advocate on behalf of GCOE researchers to develop and evolve services that address the needs of the college.

The working group noted the following improvements over time:

- Developments within Illinois Computes' resources appear to be progressing since the appointment of their director in June of 2024. Since that time, GCOE faculty have engaged successfully with Illinois Computes on Campus Cluster, Nightingale, Taiga, Research Notebooks, and other services. In time, Illinois Computes may develop to the point where a college-level

investment in Campus Cluster is no longer necessary – at least as far as the basic needs of GCOE researchers are concerned.

- Uptake in the usage of Illinois Computes resources has increased in the past year. In Spring of 2024, there were 19 research groups using Illinois Computes' queue on Campus Cluster. As of May 2025, there are 42 research groups using it. More broadly, there are 117 faculty across all GCOE departments using one or more Illinois Computes' services.

1. Introduction

At the request of Executive Associate Dean Phillippe Guebelle, a committee representing GCoE[1] faculty, GCoE administrators, and EngrIT[2] was formed to conduct a strategic-focused assessment of GCOE's research computing needs and return advisory recommendations and comments. In specific, the charge letter[3] details the following primary responsibilities:

- Reviewing policies concerning the sustainability and provision of any needed research IT that enhances the College research mission, both at the Engineering College level and at the campus level.
- Considering IT support for new research pursuits in our College, anticipating new opportunities and responding to interests and initiatives from our units.
- Monitor, assist with prioritizing, and provide feedback on the implementation of any proposals concerning research IT.
- Assist with monitoring, providing feedback, and reporting on the progress of work to align Engineering IT services with campus-level IT services, in particular with respect to research.
- Coordinating and communicating with the other IT working groups with overlapping interests.

And requests particular attention this year to these topics:

- Develop recommendations on datacenter investment within GCOE
- Explore the pros and cons of a dedicated college-level office to handle software procurement.
- Consider the current ecosystem that supports the use of AI tools for research and make recommendations for policies and investments to enhance these capabilities.
- Review and evaluate efforts related to increasing awareness of technology resources for research.
- Review the need to renew previously recurring investments in research computing GCOE has made on behalf of our researchers.
- Review any proposed investments in research computing originating from campus and provide input on how they may be best directed and utilized by GCOE faculty

[1] Grainger College of Engineering

[2] Engineering IT Shared Services

[3] See appendix for entire letter

2. Actions of the Committee

Meetings were held on: 11/25/2024, 2/20/2025, and 04/01/2025.

The committee hosted the following guests:

- Gianni Pezzarossi and Kaiwen Xue – Computational Systems Analysts on 11/25/2024 in a discussion about investment recommendations regarding datacenters.

3. Findings and Recommendations of the Committee – Specific Topics

At the request of Executive Associate Dean Philipe Geubelle, the Research Working focused on the following items, which are discussed in detail below:

3.1 Develop recommendations on datacenter investment within GCOE – 11/25/2024

The Advanced Computation Building (ACB) hosts a central datacenter available to anyone on campus. In early January 2025, ACB completed an upgrade to the redundant power system to make it more resilient to outages. In the intervening time from planning to completion of that upgrade, new concerns have arisen related to multi-GPU systems that generate more heat and consume more power than ACB can reliably provide. To that end, the AY24/25 Research Governance Working Group explored and endorsed a joint proposal from Engineering IT and Technology Services to explore the development of the National Petascale Computing Facility (NPCF) as a shared datacenter similar to ACB, but with a specific focus on hosting these advanced GPU systems.

Engineering IT helped draft a proposal which was endorsed by Research Governance and ultimately shared with both the Executive Associate Dean and the Dean of Engineering in March of 2025.

3.2 Explore the pros and cons of a dedicated college-level office to handle software procurement – 1/31/2025

Across the college, individuals in various departments face recurring challenges navigating the complex environment of software procurement for specific needs. Many software titles are purchased on one-off license agreements that require re-negotiation each time the license needs to be renewed, but staff turnover (including faculty, support staff, TA's, RA's, and project/lab managers) can add unwelcome delays and complications through a resulting lack of understanding on how to successfully complete the software purchase process. Further, with the upcoming requirements for software to meet Title II accessibility guidelines, navigating the authorization and/or exemption process for certain software will place additional burdens on individuals around the college in situations where the purchase of the software is authorized, but the use of that software in a classroom setting may not be.

Research Governance endorses the further discussion of establishing a GCOE software licensing specialist. This specialist could either be a 100% FTE or two 50% FTE's in order to ensure the needs of

the college are met in the event of the departure of this specialist. Given the relevant purchasing and procedural contacts already in place, ideally this person would be a part of the Engineering Business Services Center to ensure a focus on the needs of the college. An alternative could be that GCOE funds all or part of a member of Technology Services' Webstore to focus on GCOE needs.

Research Governance recommends the further exploration of the value this position would add to the college. By engaging with relevant departmental stakeholders, Engineering IT could help identify where this position would fit – i.e., Business Services, Technology Services' WebStore, in Engineering IT's Project Management Office, or if it should remain a departmental responsibility – and then seek approval for creating this new position.

3.3 **Discuss solutions to address the potential loss of research and other critical data contained in Box folders marked for deactivation – 1/31/2025**

This was an unplanned, but essential topic that impacted researchers across campus – not just within GCOE.

Working with Technology Services' Box storage team, Engineering IT was able to obtain a list of accounts marked for deactivation and the folders those accounts had shared with active GCOE faculty and staff. Individual lists of impacted folders were sent to those faculty and staff with instructions on how to preserve them. Engineering IT's advocacy for the preservation of this data – some of which is critical to maintain due to research agreements requiring its preservation – was significant.

Research Governance recommends ongoing efforts to emphasize key distinctions between data storage locations (both cloud and conventional), and continue to advise staff, faculty, and students on locations and strategies to both store and preserve their data long-term.

3.4 Consider the current ecosystem that supports the use of AI tools for research and make recommendations for policies and investments to enhance these capabilities – 04/01/2025

There are several campus pages dedicated to policies and practices related to the use of AI. Most notably, The Office of the Provost offers an in-depth guide to best practices for the use of generative AI in research. (https://genai.illinois.edu)

The landscape of AI tools used for research is changing at a rapid pace. For the moment, the recommendation of the Research Governance Working Group is to continue to monitor the development of AI tools, services, and policies to look for opportunities and solutions that make sense for investment on behalf of GCOE faculty. That said, monitoring developments within the National Artificial Intelligence Research Resource (NAIRR) project and the development of advice on the practicality and scalability of on-premise vs cloud resources is encouraged.

3.5 Review and evaluate efforts related to increasing awareness of technology resources for research.

Discussions by the previous Research Governance Working Group (AY23-24) included strategies to address the awareness of resources available for researchers and their teams across the college. Efforts in this area from Engineering IT have included welcome week presentations to graduate students, discussions about the IT ecosystem on campus to the GEAR program, and the publication of monthly newsletters with timely information about IT services that would impact the college.

Welcome week presentations and meeting with new faculty is an ongoing effort. The challenge with this approach is the timeliness of the information. Not everyone in the audience will be ready to act on the information shared with them in each presentation. Nevertheless, Engineering IT will continue their efforts in this regard as welcome week presents a unique opportunity to reach larger numbers of individuals new to campus.

In October of 2024, Engineering IT presented to the GEAR cohort of first year researchers on the campus IT landscape and shared information about resources that are available through Engineering IT, Technology Services, NCSA, and Illinois Computes. The attendance and responses were very positive, and it is our recommendation that Engineering IT continue to work with the GEAR program to help ensure the growth of the awareness of resources for researchers within the college.

In addition to welcome week activities and connecting with new faculty through the GEAR program, Engineering IT is increasing its efforts to engage with new faculty who have accepted offers to join GCOE. Beginning in earnest in AY23-24, this effort has evolved and refined to now include direct outreach about resources available in research, instruction, and in the administrative lives of new faculty. The goals of this are to 1) introduce new faculty to IT resources from which they would benefit and 2) to make it clear to new faculty that they have IT support available to them both prior to and during their time at GCOE. This effort is a direct result of the inclusion of dedicated research support in many startup offers, and has been extended to ensure all faculty (tenure track and non-tenure track alike) have a similar awareness of resources available to them.

Finally, beginning in September of 2024, Engineering IT started sending monthly newsletters to all faculty, staff, and graduate students. Topics included maintenance outage for the ACB datacenter that houses many GCOE systems, the end of life of the Windows 10 operating system, changes to the way we work with Box folders, and more. Read rates on these newsletters have averaged around 60% among all recipients, and we feel more information is needed to determine the effectiveness of this approach.

Continuing these efforts and approaches and exploring other potential methods of increasing resource awareness in an ongoing way is strongly recommended.

3.6 Review the need to renew previously recurring investments in research computing GCOE has made on behalf of our researchers.

Scratch Storage Service – FUNDED ($60,000)

In December of 2024, Engineering IT proposed and received approval for the creation of a service in support of GCOE faculty and staff. This scratch storage service provides more than one petabyte of raw storage to help safeguard critical research data for systems undergoing system maintenance. This initial $60,000 investment will establish an iterative flexible storage solution used by IT professionals in support of all members of GCOE. In future, it may be necessary to refresh or expand this service as the needs of the college continue to evolve.

Supplemental/Test GPU System – FUNDS REQUESTED ($20,000)

In June of 2023, a one-time funding opportunity afforded Engineering IT the ability to purchase a system with GPUs that were current at the time, and prevalent in use across the college. The purpose of this investment was to 1) provide Engineering IT a way to test software and system updates prior to releasing them broadly across the college, 2) afford researchers the ability to develop and/or evaluate code or software on similar hardware while waiting for the delivery of systems they've purchased, and 3) to afford research teams whose systems are undergoing maintenance a like-for-like computational resource to prevent their research from stopping altogether. Research teams from across the college have made use of this equipment while their system was under repair, when they needed to shift from cloud to conventional computing services, to test new code, and to test specific features of GPUs before committing to purchasing their own system(s).Given its usefulness in both a testing and supplementary capacity for GCOE researchers, the Research Governance Working Group endorses the continued investment in this service.

To that end, Engineering IT is requesting a modest $20,000 investment in this service to address a current service gap in support of the latest Blackwell generation of Nvidia GPUs as well as AMD's EPYC Genoa CPUs to help ensure hardware purchased by GCOE researchers over the next 18 months can be adequately supported. Additionally, the college should prepare to further invest in this service in mid-2026 when Nvidia's next generation GPU (codenamed 'Rubin') is expected to come to market.

GCOE Campus Cluster Investment – FUNDS REQUESTED ($275,000)

At present, the most pressing issue we see is that the A10 GPUs that make up the bulk of the GCOE Campus Cluster investment are not powerful enough for research work in AI/ML. These GPUs were part of an investment made in the AY22/23 cycle and were adequate for their intended purpose at the time. While advancements in GPU technology in the intervening years has limited their usefulness for larger research purposes, they remain suitable for myriad instructional use cases. The primary issue we face in the instructional space is that we don't have enough GPUs for a large class to provide each student a dedicated GPU for the entire semester. If we begin transitioning our A10 nodes in the research queue over to the instruction queue, and invest in a node with multiple H200 GPUs, that new node would be available for use as either a single node with 8 H200s, or it could be subdivided into into as many as 56 smaller GPUs depending on the specific needs of researchers or instructors in their course projects.

3.7 Review any proposed investments in research computing originating from campus and provide input on how they may be best directed and utilized by GCOE faculty

Illinois Computes has made numerous investments and adaptations in the past year that have

improved the capabilities of the services in their catalog. Increasing long-term storage allocations from 10TB to 50TB, the addition of 8 Nvidia H200 GPUs in the Campus Cluster, and working to increase the functionality of the Illinois Computes Research Notebook service with 11 A100 GPUs are notable enhancements to these services.

Through their relationship with the Director of Illinois Computes as well as the different service managers in charge of the myriad services offered by Illinois Computes, Engineering IT is proactively advocating for the increased usage of those services and ensuring that GCOE's faculty needs are met in terms of service delivery and enhancement.

The Research Governance Working Group recommends that Engineering IT and Illinois Computes continue to work together in this capacity to further strengthen their relationships with GCOE faculty and to address their needs as they continue to evolve.

4. Findings and Recommendations of the Committee – Additional Topics

The committee also discussed and has recommendations/comments in the following additional areas that fall within the general charge of the committee.

4.1 Improvements Noted

Developments within Illinois Computes' resources appear to be progressing since the appointment of their director in June of 2024. Since that time, GCOE faculty have engaged successfully with Illinois Computes on Campus Cluster, Nightingale, Taiga, Research Notebooks, and other services. In time, Illinois Computes may develop to the point where a college-level investment in Campus Cluster is no longer necessary – at least as far as the basic needs of GCOE researchers are concerned.

Appendix 1: Charge Letter

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
THE GRAINGER COLLEGE OF ENGINEERING

306 Engineering Hall MC 266
1308 West Green Street
Urbana, Illinois 61801

Philippe H. Geubelle
Bliss Professor of Aerospace Engineering
Executive Associate Dean

(217) 244-7648
geubelle@illinois.edu
http://ae.illinois.edu/directory/profile/geubelle

October 10, 2024

Andre Schleife (schleife@illinois.edu, MatSE) - Chair
Nick Colaneri (nfc@illinois.edu, GCOE office of Research)
Mark Hart (mshart@illinois.edu, Engineering IT – Ex Officio)
Mark Neubauer (msn@illinois.edu, Physics)
Jianqi Xi (jianqixi@illinois.edu, NPRE)
Jinhui Yan (yjh@illinois.edu, CEE)

Dear Colleagues:

Thank you for agreeing to serve on the Grainger College of Engineering IT Governance Research Working Group for the academic year 2024-2025. This working group is critical to helping ensure that IT support of the research needs of faculty, researchers, and graduate students in the College is of the highest quality and reliability.

Your recommendations and comments are advisory to the Executive Associate Dean, with primary responsibilities for:

- Reviewing policies concerning the sustainability and provision of any needed research IT that enhances the College research mission, both at the Engineering College level and at the campus level.
- Considering IT support for new research pursuits in our College, anticipating new opportunities and responding to interests and initiatives from our units.
- Monitor, assist with prioritizing, and provide feedback on the implementation of any proposals concerning research IT.
- Assist with monitoring, providing feedback, and reporting on the progress of work to align Engineering IT services with campus-level IT services, in particular with respect to research.
- Coordinating and communicating with the other IT working groups with overlapping interests.

In particular, over this academic year, I would like to ask you to

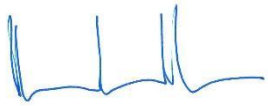
- Develop recommendations on DataCenter investment within GCOE

- Explore the pros and cons of a dedicated college-level office to handle software procurement.
- Consider the current ecosystem that supports the use of AI tools for research and make recommendations for policies and investments to enhance these capabilities.
- Review and evaluate efforts related to increasing awareness of technology resources for research.
- Review the need to renew previously recurring investments in research computing GCOE has made on behalf of our researchers.
- Review any proposed investments in research computing originating from campus and provide input on how they may be best directed and utilized by GCOE faculty

You will also be asked to provide an assessment at the end of the academic year (no later than June 1, 2025) on the current state of IT support for research activities. Dr. Schleife has graciously agreed to chair this working group. He will be in contact with you soon to arrange for your first meeting.

I am thankful to all of you for your willingness to contribute to this important work for our College.

Sincerely,



Philippe H. Geubelle

Appendix 2: GCOE Campus Cluster Usage Statistics

Research Queue Usage 1/2024 - 4/30/2025

Partition (Queue) Name: eng-research-gpu
Total number of jobs started: 112,152 Jobs
Total CPU hours consumed: 407,620 Core-Hours
Total GPU Hours Consumed: 65,161 GPU-Hours
Average GPU Job Size: 1.01 GPUs
Average Wait Time: 0.47 Hours
Unique users: 67

Instructional Queue Usage 1/2024 - 4/30/2025

Partition (Queue) Name: eng-instruction
Total number of jobs started: 23065 Jobs
Total CPU hours consumed: 262,890 Core-Hours
Total GPU Hours Consumed: 474 GPU-Hours
Average Wait Time: 3.89 Hours
Unique users: 133

Depts. Using the Instructional Queue

Siebel School of Computing & Data Science
Nuclear, Plasma, & Radiological Engineering
Electrical & Computer Engineering
Information Sciences
Statistics
Astronomy
Physics
Civil & Environmental Engineering
Materials Science & Engineering
Linguistics (Siebel School of Computing & Data Science)

Appendix 3: Meeting notes

https://docs.google.com/document/d/1Es8AWPEi5-GakLKZEGbfYdU9TkqlNsWY_dc6Q_nq1JM/edit?tab=t.0#heading=h.9e9i8upysskt