

Provost's Strategic Task Force on Cyberinfrastructure and Data Science

Report & Recommendations

June 15, 2023

Mark Aakhus, Alyson Brooks, Yuebin Guo, Ronald Hart, Regina Liu, Meredith McGill,
Thu Nguyen, Katherine Ognyanova, and Jaideep Vaidya

Taskforce Members

Mark Aakhus, Co-Chair, Communication, Dean's Office School of Communication and Information
Barr von Oehsen, Co-Chair, Office of Advanced Research Computing
Alyson Brooks, Physics, School of Arts & Sciences
Jaideep Vaidya, Management Science & Information Systems, Rutgers Business School
Katherine Ognyanova, Communication, School of Communication and Information
Meredith McGill, English, School of Arts & Sciences
Regina Liu, Statistics, School of Arts & Sciences
Ronald Hart, Cell Biology & Neuroscience, School of Arts & Sciences
Thu Nguyen, Computer Science, Dean's Office School of Arts & Sciences
WooJin Jung, School of Social Work
Yuebin Guo, Mechanical and Aerospace Engineering, School of Engineering

Major Stakeholders

Yana Bromberg, Biochemistry and Microbiology, School of Environmental and Biological Sciences
Stephen Burley, Chemistry and Chemical Biology, School of Arts and Sciences and the Institute for
Quantitative Biomedicine
Enrique Curchister, Environmental Sciences, School of Environmental and Biological Sciences
Richard Ebright, Chemistry and Chemical Biology, School of Arts and Sciences and the Waksman
Institute of Microbiology
Lauren Goodlad, English, School of Arts and Sciences
Shatenu Jha, Electrical & Computer Engineering, School of Engineering
Matt Weber, Communication, School of Communication and Information
John Wilkin, Marine and Coastal Sciences, School of Environmental and Biological Sciences

Table of Contents

<i>Executive Summary</i>	3
<i>Recommendation: Create a Chancellor Level Institute for AI, DS, CI</i>	4
The Problem & Solution.....	5
The Strategy	6
The Method.....	7
The Projected Goals	9
The Ask	10
The Value.....	10
<i>Recommendation: Advance Rutgers Research Cyberinfrastructure Ecosystem</i>	11
Critical Issues for Advancing RU’s Research Cyberinfrastructure	11

Executive Summary

To ensure that RU-NB remains a top-tier research institution, we must integrate, coordinate, and expand RU-NB's intellectual strengths around the areas of artificial intelligence (AI), data science (DS), and cyberinfrastructure (CI). We must do so in ways that advance knowledge and translate innovations to a broad range of fields of research, scholarship, and creative activity to maximize their scholarly and societal impact. The need for a coherent, comprehensive strategy to achieve these goals is profound.

To seize this moment, the members of the *Chancellor-Provost's Strategic Taskforce on Cyberinfrastructure & Data Science* investigated the intellectual, institutional, and technological landscape at RU-NB and its peers. Our primary findings include:

- The consequences of social and technical changes related to AI, DS, and CI are significant, and in some cases 'existential threats' according to an RU-NB expert, for nearly all branches of creative inquiry including the arts, humanities, sciences, engineering, and professional schools.
- Ongoing developments around AI, DS, and CI are changing the conditions for disciplinary, multi-disciplinary, interdisciplinary, and transdisciplinary knowledge production.
- Rutgers students require these approaches for research competitiveness and career development.
- While RU-NB has faculty and staff strengths in these areas, it is currently difficult to identify or leverage overlapping strengths to advance research, education, and outreach.
- RU-NB requires new capabilities for coordinating the collaborative development of and critical engagement with AI, DS, and CI to advance research, scholarship, and creative activity.

Remaining a top-tier research institution will require access to ultramodern research cyberinfrastructure and accompanying software expertise and support that accelerates faculty and student engagement in world class research, education, and outreach. It also requires new organizational pathways that enable faculty and students to easily work together across their disciplinary fields and the organization's internal boundaries. It is incumbent on university administrative leadership to facilitate the institutional clarity necessary to realize such a strategy, especially the flow of funds, design of services, and policy framework that ensures the university structure serves one of its primary functions of discovery and innovation. This is an enormous and crucial task we must all embrace.

Our recommendations:

- **Create a Chancellor level Institute at RU-NB for AI, DS, CI** to serve as a crucial campus catalyst for digitally enabled, data-driven discovery within and across domains that brings faculty, staff, and students together to address grand scientific and societal challenges related to AI, DS, and CI.
- **Advance RU's Research Cyberinfrastructure Ecosystem** to enable achievement of system-wide goals for academic excellence by building from prior commitments to advanced research computing with a renewed strategy that leverages centralization and federation of technical-services and end user support for research, scholarship, and creative activity.

Our recommendations are grounded in our fundamental commitment as scholars to advance knowledge by constructing the best possible enabling conditions for knowledge production. Such a commitment embraces open scholarship enabled by diversity, equity, inclusion, accessibility, and environmental sustainability. These recommendations are crucial starting points for a deeper campus engagement for their strategic development and implementation.

Recommendation: Create a Chancellor Level Institute for AI, DS, CI

As New Jersey's premier public university—one that is recognized in the Carnegie classification as a research-intensive university and a member of the Big10 Academic Alliance—Rutgers New Brunswick holds an elite and privileged position that comes with an obligation to deliver field-defining research accessible to all for the benefit of New Jersey and beyond. RU-NB cannot maintain let alone advance this position without a serious intellectual and material commitment to an integrated strategy at the nexus of Artificial Intelligence (AI), Data Science (DS), and Cyberinfrastructure (CI).¹ Advances in these areas are significant for all branches of research, scholarship, and creative activity.

Research universities like Rutgers need to find new ways to address fast-paced innovation in these fields, which advance our capacity to manipulate the fundamental processes of life and to engineer radically new built environments. Such developments promise to transform society and culture, while urgently calling for our participation in the discourses through which we understand and shape our world. For the benefit of society, research universities need to reinvent their approaches to producing breakthroughs in knowledge. Developments in AI, DS, and CI call for new intellectual, social, cultural, and technical capabilities (see [Appendix A](#)). How can RU-NB approach these challenges with imagination, to the advantage of Rutgers, the State of New Jersey, and beyond?

It is clear from our assessment that RU-NB is positioned for leading roles in AI, DS, and CI, including critical engagement with the social, cultural, and environmental implications of such innovation. Doing so will depend on three fundamental infrastructural factors underpinning the community of research and scholarship. The first is technical: we need dependable state-of-the-art means for research, pedagogy, and outreach with the techniques of and within the fields of AI, DS, and CI. The second socio-technical: our technological systems must be embedded within a vibrant intellectual ecology that affords our faculty, staff, and students collaborative platforms for re-thinking approaches, methods, and models foundational to their research, scholarship, and creative activity. The third is more organizational-cultural: there is no question that RU-NB already has significant talent and

¹ We define these overlapping fields of inquiry as follows:

- Artificial Intelligence (AI) is an area of research focused on the development of computing to perform tasks that involve some dimension of reasoning, generalizing, discovering meaning, learning from experience, or generating content. AI includes machine learning (ML), deep learning (DL), natural language processing (NLP), large language models (LLM), and robotics. ML focuses on developing statistical tools for exploring data through approaches such as supervised, unsupervised, and reinforcement learning. DL focuses on developing machine learning methods based on artificial neural networks inspired by biological systems. NLP focuses on technologies that allow machines to understand, interpret, and generate human language. LLMs are ML models trained on massive amounts of text data for text generation and language-based tasks. Robotics focuses on the design, construction, operation, and use machines that can be programmed to perform a range of tasks.
- Data Science (DS) is an interdisciplinary research field focused on effective ways to develop insights from data. DS draws on AI, ML, and DL as well as statistics, mathematics, computer science, and informatics to understand problems, collect, clean, and explore data, build models, and develop insights. Fair and Responsible Data Science (FRDS) incorporates methods from across the social and behavioral sciences and the humanities to mitigate unwarranted bias in DS pipelines.
- Cyberinfrastructure (CI) research develops technological, sociological, and communication solutions for organizing and connecting laboratories, data, computing, software, and people into environments that can produce new knowledge, theoretical breakthroughs, and novel solutions to research challenges.

accomplishments across AI, DS, and CI, but these strengths remain diffused and siloed in ways that do not leverage or accelerate that potential for a vibrant intellectual ecology, one that benefits our students, our staff, our faculty, and our stakeholders.

Conversations with our colleagues and peers at other institutions have made it clear that RU-NB needs an integrated strategy to organize and elevate thought leadership from our scholarly community. The Institute we outline in this recommendation would define and implement pathways for new research, educational programming, and stakeholder outreach. Such an institute would attract exciting new scholars, students, and community and industry partners into a vibrant intellectual ecology around AI, DS, and CI, and would play a vital role in advancing research cyberinfrastructure at RU-NB, attracting regional, national, and international attention, partners, and funding. The Institute we outline would achieve critical goals outlined in the RU-NB Academic Master Plan that are also necessary for the success of related Chancellor-Provost Strategic Taskforces on Climate Change and Life Sciences.

The Problem & Solution

We propose the creation of an Institute at the Chancellor-Provost level to serve as a catalyst to foster creative achievements in research and education by breaking down academic silos, enabling collaboration, and attracting external funding (see [Appendix B](#)). The Institute's purpose is to transform the current situation at RU-NB, in which a thicket of activity relevant to AI, DS, and CI proliferates—often unsystematically—into a vibrant, coherent intellectual ecology organized around AI, DS, and CI.

RU-NB undoubtedly has considerable research strengths in AI, DS, and CI but its current circumstances would be transformed by the timely intervention we propose (see Appendices [C](#) and [D](#)). The current environment is characterized by disjointed efforts, misaligned or unavailable resources, untapped opportunities, and an urgent need to stabilize basic conditions for productive work. The overwhelming majority of our colleagues would like a different experience—one marked by clear institutional guidelines, opportunities to connect with peers, and structures that promote discovery for individual and collaborative research.

The solution we offer requires that we identify and nurture areas of opportunity for discovery. As characterized in Figure 1, there are real-world and disciplinary problems in urgent need of AI, DS, and CI methods. Likewise, progress and breakthroughs in the fields of AI, DS, and CI often happen when they are applied to knotty real-world and discipline-specific problems. The fundamental role of the Institute is to define the pathways that will bring problem-solvers with different kinds of expertise together and, more generally, to enable a shift of research structure and culture at RU-NB. Drawing insights from our peers (see [Appendix E](#)), we anticipate that providing a structure for integrating AI, DS, and CI with domain expertise across the sciences, social sciences, and humanities at RU-NB will galvanize scholarship at the local level and advance science and scholarship of national and international significance.

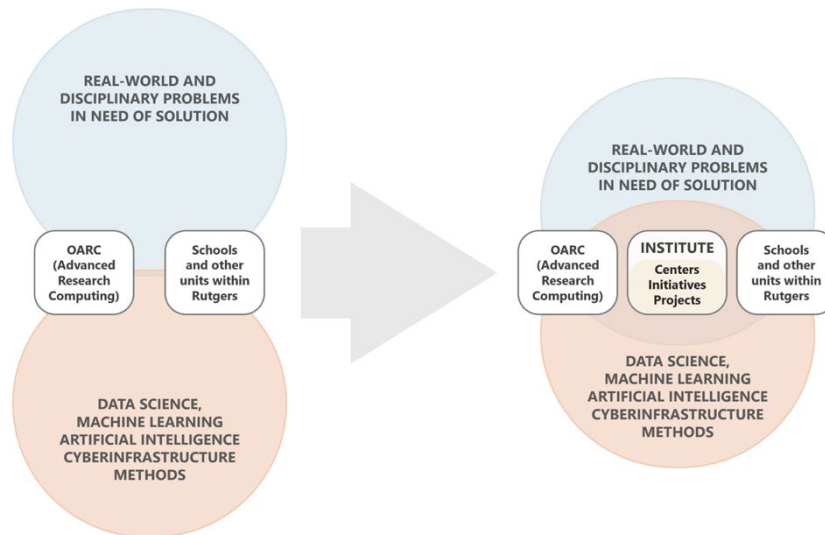


Figure 1: Transforming RU-NB's AI, DS, and CI research context.

To achieve this transformative goal, we recommend that this Institute be created with the following purpose and mission:

To use AI, DS, and CI to incubate and accelerate domain-specific advances and multi-disciplinary solutions to grand scientific and societal challenges. The Institute will foster AI, DS, and CI research at RU-NB to explore its translational impacts for science, scholarship, workforce development, and community engagement while striving to create more inclusive, transformational forms of inquiry.

The Strategy

The Institute will advance research at RU-NB along two primary axes:

First, the Institute's main activities will advance AI, DS, and CI research while facilitating its translation in and between various domains of inquiry. We highlight below the Institute's method in bringing faculty, students, and industry and community partners together to push the frontiers of knowledge.

Second, the Institute will partner in the innovation of research cyberinfrastructure (broadly conceived as an ecosystem) and in ensuring high-reliability, state-of-the-art research cyberinfrastructure services (also broadly conceived). In achieving its fundamental academic mission, the Institute will play a decisive role in elevating RU-NB's organizational capabilities for digitally enabled research. Doing so locally at RU-NB will define pathways for discovery and innovation key to unleashing our potential as a research university while also identifying RU-NB as a leader in integrated intellectual and administrative approaches to AI, DS, and CI. The Institute's success depends upon a robust research cyberinfrastructure, which is addressed in part two of our recommendation.

The Institute's strategy is summarized in Figure 2. The activities depicted on the left involve coordinating, cooperating, and partnering with Schools, Institutes, Centers, and other existing relevant academic activities. The



activities on the right involve partnerships with core University services, especially OARC, but also the Office of Information Technology, the Office for Research, and the New Brunswick Libraries.

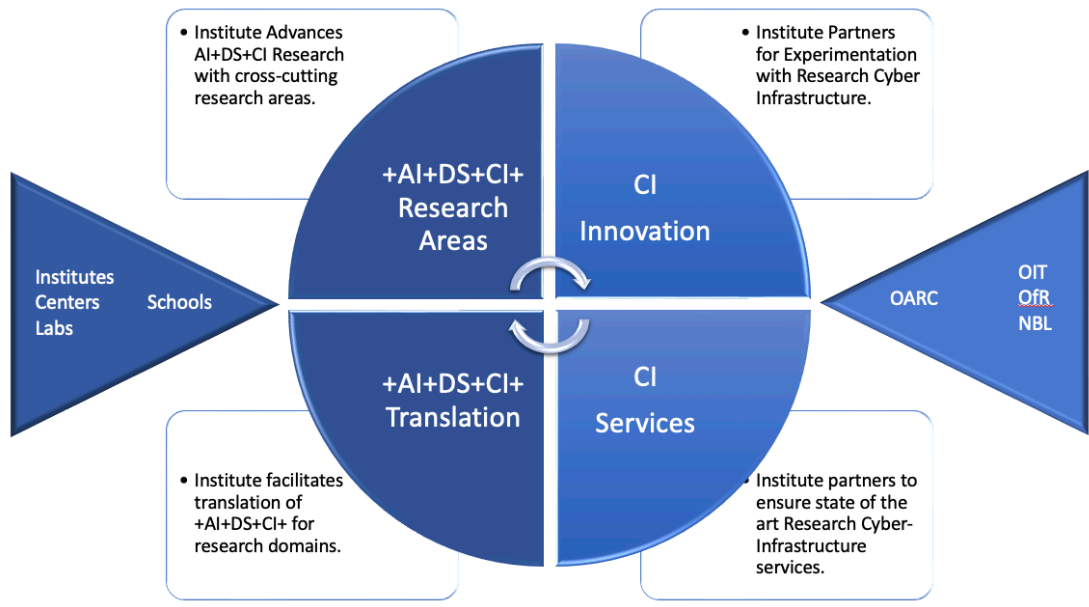


Figure 2: Institute Strategy for Research and Research Cyberinfrastructure

The Method

The Institute will function as an incubator and accelerator of discovery and innovation (see Fig. 3). The principal focus will be the startup, development, and lifecycle of research projects. These projects will focus on the breakthroughs that can be made where research in DS, AI, and CI intersect with each other and especially with domain-specific research, scholarship, and creative activity at RU-NB.

Projects will be developed within the Institute and in partnership with other academic and administrative units at Rutgers as well as with external community, industry, and governmental stakeholders. As warranted by their success and continued relevance, projects may become established as initiatives and centers within the Institute or adopted within other RU-NB academic and administrative units or with other external stakeholders. This method emphasizes organic development of projects with respect for their lifecycle. We see two fundamental sources for projects:

First, the Institute will focus on coordinating and elevating existing research strengths while defining new areas of discovery and innovation. As schematized in Figure 3, we envision the Institute as defining and refining reliable methods—research innovation pipelines—that will enable investigators at all career stages to come together to solve real-world and disciplinary problems that require AI, DS, and CI to address. Innovation pipelines identify and give shape to “sweet-spots” for discovery that might originate from any of the domains of research, scholarship, and creative activity at RU-NB, or within the fields of AI, DS, CI themselves.

The Institute will experiment with tools and practices and with the organization of research to create effective innovation pipelines, such as: a web-based platform for finding collaborators and projects, open innovation platforms, seed-funding programs, clinics, ideation-to-submission support, cluster hiring, speaker series, summer schools, agile development methodologies and DevOps for digitally enabled science and scholarship, training

workshops, communities of practice for doctoral students, experiential learning programs for undergraduates and masters students, and so on. The projects that emerge may result in research activity, educational programs, outreach initiatives, or some combination of these, but all will converge to advance a culture of discovery and innovation around AI, DS, and CI.

In this way, the Institute will build a portfolio of activities that excites the research and educational community at RU-NB, attract external funding, and engage external public and private partners within New Jersey and beyond.

Second, the Institute will focus on generating a matrix of opportunities for organizational learning about building and sustaining agile and high-reliability services for vibrant research environments. These opportunities will primarily arise around needs for data management and storage, high performance computing, collaboration and sharing, access to tools and frameworks, and training. Recognizing that the tools and technologies for conducting research, scholarship, and creative activity are a fundamental condition for success, the Institute will play a crucial role in the evolution of RU-NB's research cyberinfrastructure ecosystem.

The Institute will coordinate and partner with OARC, RU-NB Libraries, and other core services to foster agile development and continuous integration/continuous deployment. Together they will define necessary developments of production level services while partnering to experiment with the development of novel approaches that advance the research cyberinfrastructure. These activities will require specialists who can facilitate key aspects of digitally enabled research such as data acquisition and management, data processing and cleaning, feature engineering, model validation, and visualization. This focus on research infrastructure will in turn open significant possibilities for RU-NB students to participate in novel experiential learning programs (e.g., research credits, internships, co-ops, fellowships), programs keyed to newly emerging roles in the workforce produced by advances in AI and computing.

Capitalizing on these opportunities, the Institute will become a vital hub for advancing the research cyberinfrastructure ecosystem (locally, regionally, nationally, and globally), especially the development of data science and cyberinfrastructure professionals. This activity will excite the RU-NB research community, attract external funding to advance RU-NB's cyberinfrastructure, and engage external partners.

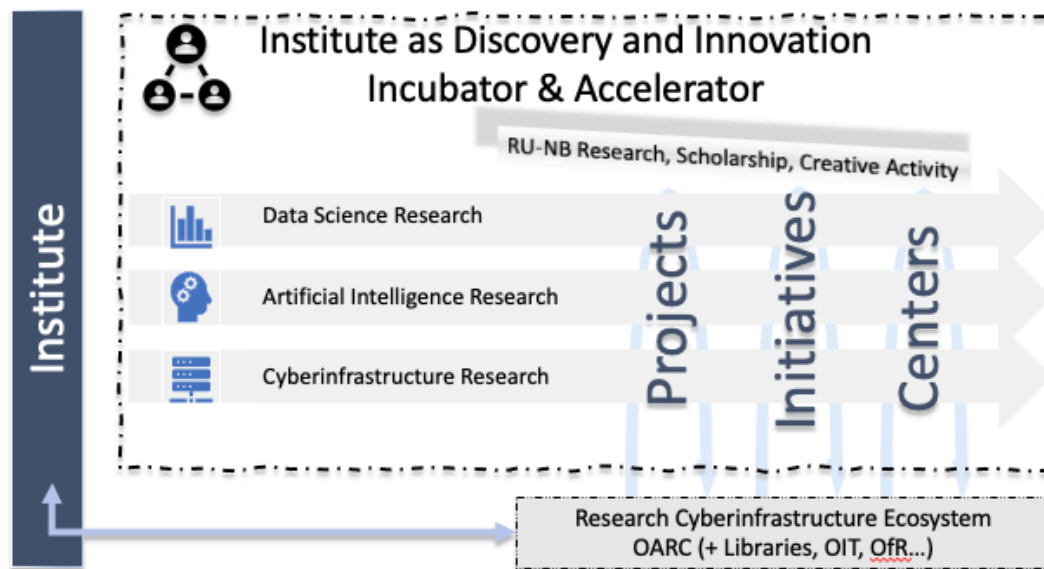


Figure 3: Institute's Basic Function

The Projected Goals

The key opportunities the institute can leverage at RU-NB are twofold:

First, the institute will facilitate the development of projects that address grand challenges related to AI, DS, and CI that cut across disciplinary areas. Some areas where RU-NB has considerable strengths suggest potentially viable starting points for collaboration that include, for instance:

- Trustworthy and Safe AI
- Social Inequality and Democratic Participation
- Computational Modeling and AI for Science
- Future of Work and Economic Development

Second, as discovery and innovation incubator, the institute will engage our research community in developing digitally enabled, data-driven discovery within and across domains. Such activity builds strength within and across domains while scaffolding deeper inquiry into digitally enabled research and scholarship and its translation into practice. Some critical starting points include, for example:

- Bio/Life Sciences: The development of novel algorithms and strategies for data analysis across the variety of biological data, including genomic investigation of disease using high-throughput sequencing and single-cell technologies.
- Physical and Environmental Sciences: The development of AI-enabled science for working in with big data in peta and exa scale computing environments to model the Universe, oceans, climate, weather, and materials.
- Engineering: Incorporating engineering principles to advance machine learning algorithms for learning from sparse or unstructured data, domain knowledge, and side information often encountered in real-world applications.

- Social Sciences: Advancing computational social science for understanding social processes in emerging information environments.
- Humanities: Spur development of techniques for reading at scale, interpreting digitized images, and data visualization for unstructured data and critical engagement with AI.
- Community engaged research: Clinics that bring together domain scientists, statistics/computer science/electrical and computer engineering researchers, and research specialists in the libraries and OARC with experts in community engagement and policy to define and advocate for modern engaged scholarship within the university and beyond.

The Ask

Our recommendation will require an investment from RU-NB for the Institute to be established and for its success to be sustained. These requirements will include:

- Personnel: Director, Faculty Research Area Leads, Research Engineers and Facilitators, Staff, Post-docs, Graduate Students, Fellows, Interns, and hourly workers.
- Programs: Seed Funds, Cluster Hires, Program Development Funds, Events Funds.
- Space: Offices, Meeting Space, Labs
- Technology: Scientific and Research equipment, Computer hardware and software, Web Services, Networking and Connectivity, Administrative tools

The Value

The Institute will realize important high-level outcomes for RU-NB's capabilities while reinforcing the core values of a public, research university for the 21st century. These include:

- Synergistic academic activity involving faculty, students, and the community.
- Progress and breakthroughs in research, scholarship, and creative activity.
- New and novel programs for education, training, and outreach related to AI, DS, and CI.
- A recognized hub that demonstrates and communicates RU-NB leadership both internally and externally.
- Organic and stable life cycle of projects, initiatives, and centers that appeal to faculty, students, and external investors.
- Increased external funding for projects, initiatives, centers, and research cyberinfrastructure from federal agencies, foundations, the State, and industry.
- Internal innovations in providing services and support for investigators across a range of domain interests, including AI, DS, and CI (e.g., informatics, software development, data management, FAIR principles, democratizing access).

The charge given to the CI-DS Taskforce is an urgent one. It calls for an intervention into fundamental infrastructural issues that have limited research and instruction at RU-NB. While the technical challenges and opportunities introduced by AI-DS-CI are significant, the organizational-cultural issues on which they rest are no less critical to the success of the next generation of research at RU-NB. The Institute we propose is a down-payment for organizational change that invests in a strategy and method for building a strong foundation for discovery to flourish. In [Appendix B](#), we provide further detailed suggestions about the mission, programming, structure, governance and financing of the Institute.

Recommendation: Advance Rutgers Research Cyberinfrastructure Ecosystem

For RU-NB to be a premier research university, a state-of-the-art research cyberinfrastructure ecosystem is mission critical.² The initial commitment to advanced research computing at RU yielded significant advances, especially in system-wide equity and in terms of democratizing access to resources (see appendices [G](#) and [H](#)). However, while providing a good start, these initiatives need to be further refined and developed to achieve the system-wide goal of academic excellence. For Rutgers, it is critical to identify the next-level advance in research cyberinfrastructure for science, scholarship, and creative activity, which will require bold, visionary leadership.

A large proportion of current research depends on computing and data services, with data intensive research and artificial intelligence becoming leading drivers. Thus, a robust research cyberinfrastructure ecosystem is absolutely critical to knowledge production and innovation. University-wide research computing and data enables affordability and access to research resources that both saves departments and schools money while facilitating more grant funding and innovation. Investing in the research cyberinfrastructure ecosystem, which includes its workforce, increases a university's competitiveness for external funding, faculty recruitment, publications, and research with global impact.

RU-NB, like the other campuses, has a critical stake in how RU's research cyberinfrastructure ecosystem is maintained and developed. Our taskforce has found that RU is at a critical turning point to remain relevant with trends in research and competitive with its peers. In particular, RU-NB loses excellent faculty to universities with better infrastructure and services while current limitations continue to undermine our investigator's competitiveness for external funding.

Our recommendation is that RU build from its current strategy for advanced research computing to a strategy for a robust comprehensive research cyberinfrastructure ecosystem. We recommend a strategy that includes a mix of centralized and federated computing resources and services along with external cloud infrastructure. Doing so will enable RU to achieve the system wide goal of academic excellence. We believe our colleagues across Rutgers will share our perspective along with additional interests unique to their campuses.

Critical Issues for Advancing RU's Research Cyberinfrastructure

Going forward, there are several possibilities in terms of how computing and data resources are provided. Computation and data resources can be centralized, accessible through the cloud, tightly federated, or loosely federated. We propose the following key issues to be addressed:

1. Focus on OARC's core mission to limit mission creep: The key mission of OARC is to provide seamless institution-wide access to computing and data that is transparent in terms of the location and modality of computing resources.

² See Appendix F for definition of research cyberinfrastructure ecosystem elements based on the National Science Foundation's conceptualization: (1) *the people*, organizational units, and communities involved in research, scholarship, and creative activity; (2) *coordination and user support* (allocation services, end-user support services, operations and integration services, monitoring & measurement services, technology translation services), and (3) *technical-service layers* (computing resources and services, data infrastructure, gateways & hubs services, public access & open science, pilots & testbeds, software & workflow systems, CI-enabled instrumentation, research & education networks, and security layers).

2. Develop and deploy a financially sustainable cloud model. The cloud has several key advantages in terms of providing high availability, elasticity, scalability, and cost effectiveness. However, indirect costs are currently applied to all cloud expenses, making it financially uncompetitive, even though using the cloud does not require any additional management burden on the university. Funding agencies such as NSF have recognized this and are currently making alternative arrangements such as providing direct funding for CloudBank (Funds that go through CloudBank are free of indirect costs, maximizing the value of NSF awards). However, this also means that the funds never come to the University. We recommend that the University revisit this issue and make sure that cloud expenses are not unduly taxed and thus disincentivized.
3. Prioritize OARC's transparency in services. If data is within a centralized condominium such as Amarel, on the cloud, or within local clusters managed within a unit, it should be transparently accessible. OARC should provide expert advice regarding the most appropriate model for computation/storage and focus efforts to make access easier and integrated.
4. Incentivize faculty and units to purchase and coordinate infrastructure and services through OARC, to the extent possible, recognizing that there may be specific situations where local ownership is necessary. This requires reducing friction between provision and acquisition services through OARC and appropriate service level agreements that ensure faculty and units are confident their requests will be effectively met.
5. Establish data protocols and repositories for open science and scholarship architectures with privacy, security, and FAIR+ principles that meet federal mandates. This can be a service provided in partnership with the RU Libraries.
6. Address the financial sustainability of computing and data resources by acknowledging its total cost. The fiscal sustainability of research computing is extremely important, and many of the hidden costs are not currently exposed to faculty or to units (e.g., Amarel condominium model).
7. Address the financial sustainability of computing and data resources by developing hybrid models for compute and data services with other units that are fiscally sustainable. RU-NB could lead by example by increasing its financial support of OARC for services to RU-NB investigators.
8. Develop an effective governance model. Build on the current faculty advisory group to include a diverse range of user communities to review and develop the service catalogue. Establish a mechanism to ensure mission and fiscal accountability by research deans and executive deans.
9. Develop new approaches for research consulting, training, and programmer support. These services are critical from the perspective of data science yet may not be central to the mission of OARC itself. The recommended Institute in part 1 is envisioned as a synergistic partner. It may even be a better home for the research scientists and the educational activities. It may also offer a more fiscally sustainable model and a vehicle to better operationalize with other units these fundamental elements of a robust research cyberinfrastructure ecosystem.

Overall, a strategic view of RU's research cyberinfrastructure ecosystem needs to be taken, and OARC itself must be viewed in this broader context along with the Libraries and the Institute we have recommended to achieve the research goals for Rutgers.