

Rutgers University: Office of Advanced Research Computing (OARC) Facilities, Equipment, and Other Resources

The Rutgers Office of Advanced Research Computing (OARC) is a centrally funded support organization providing technical and scientific support and training to faculty, staff, and students in the use of its extensive computing, networking, storage, and software resources. OARC resources available at no cost to the university research community include a combination of leading-edge technology and research support expertise.

OARC's team of computational scientists are experienced independent researchers with deep expertise in biochemistry, biomedical engineering, chemistry, computer science, data science, image processing, computational genomics, geospatial information systems, mathematics, microbiology, molecular modeling, mechanical engineering, physics, and many other fields.

The Rutgers University Platform Research Systems infrastructure team provides expertise in storage management, security, advanced networking, and automation. Together, and in collaboration with OARC's other specialized roles, these technical teams support and collaborate with investigators on demanding problems requiring advanced cyberinfrastructure to facilitate a path to solution.

Primary Computing Systems

The majority of OARC's resources are managed under a federated community cluster model named *Amarel* in memory of the prominent Rutgers artificial intelligence researcher, Dr. Saul Amarel. This, OARC's primary computing system, provides free, open access use to all students, staff, and faculty within the Rutgers research community as well as immediate, priority access for investigators who purchase compute nodes. Owners purchase the dedicated resources they need (compute nodes and storage) and OARC, through its university funding, supplies the matching infrastructure (e.g., power, cooling, network access, security, administration) along with both technical and scientific support. Jobs submitted to OARC's computing systems are managed in multiple ways depending on job type (owner or open access), resources requested, current utilization of federated cluster members, and a variety of other factors. OARC's systems are also capable of enabling bursting of submitted jobs into commercial and academic cloud services when local resources are heavily utilized, or other factors favor external resources.

The Amarel community cluster is the umbrella for distributed resources that number approximately 770 nodes providing 37,468 compute cores, 242 NVIDIA GPUs (Tesla, Volta, Titan, Ampere, Ada Lovelace, Nvidia Data Center/Tesla, and Hopper models), Intel Arria FPGAs, and Intel N3000 NICs. Several memory configurations are supported with capacities up to 2TB and including Intel Optane persistent memory modules. All cluster nodes are connected with Mellanox InfiniBand FDR, EDR, HDR and NDR. Major components of this distributed infrastructure are installed at the Piscataway, Camden, and Newark campuses. Resources are built out through funding from university, state, and federal agencies.

OARC's main storage facilities include four IBM Spectrum Scale appliances distributed across the University with an aggregate of approximately 10 PB of usable capacity. Centrally-managed NAS storage provides backup and archive capacity for OARC's systems on the various campuses. To facilitate distributed data services, Spectrum Scale AFM (Active File Management) is used to provide cached copies of user account data across federation members. AFM utilizes OARC's 100 Gbps Rutgers Research Network, initially named "CICnet" after the NSF-funded award that provided partial kick-off funding, and described in more detail in the next section. AFM utilizes dedicated data transfer nodes (DTNs) architected to optimally utilize the capabilities of the Rutgers Research Network and have been designed to move data across the network at speeds approaching 100 Gbps. The FIONA recipe for DTN has been also adopted by OARC to provide an affordable but flexible high-performance path to allow departments, labs, and other communities with demanding data movement requirements a simplified path to aggregate local data and move it into the OARC-federated compute environment.

The Rutgers Clinical and Research Data Warehouse (CRDW), housed in the Administrative Services Building Data Center (ASB) in Piscataway, New Jersey, is a HIPAA compliant centrally supported data storage system that is used to host and manage identifiable data. The data loaded into the CRDW is transmitted via secure means (e.g. VPN tunnels, Secure FTP, etc) from the source systems. The source systems include but are not limited to: Clinical EMR Data; Genomics data sets; Pathology Records; Claims and Payment data sets. The raw data that has not been processed for use in the research project will be maintained in a secure database on servers utilizing AES-128. The infrastructure meets the standards needed to meet the FIPS 140-2 requirements mandated by most government grants. A secure Extract-Transform-Load (ETL) process is in place to provide controlled access to sensitive data and convert them to de-identified datasets when required. Honest Brokers are used to allow for the re-linking of the de-identified data should an approved study need that ability if needed. Data access will be through a secure environment using Citrix and PowerBI desktops, using two factor authentication and secure encrypted links (over encrypted SQLNet using "RC4_128" or "3DES168" encryption types), limiting access to data sets specifically approved by IRB/SRB. The data security measures outlined above mitigate the possibility for breaches of patient identities.

The OARC Data Analytics Protected Environment (Amarel-PE) is designed for those research projects that require access to identifiable data. Housed in the Rutgers IBM Modular Data Center in Piscataway, the Amarel-PE solution provides the medical research community access to a research computing and data infrastructure (bare metal, virtualization, networking, and storage) designed specifically to work in tandem with the Clinical and Research Data Warehouse (CRDW). The Amarel-PE security features include data encryption, network segmentation, management of client authentication, service discovery, distributed multi-tenant authorization, and multi-factor authentication. Other security measures include standard Linux features such as firewalls, port filtering, encrypted logins, kernel/user space separation, user management, and monitoring of syslog files. Besides virtualization, the Amarel-PE architecture

includes Graphics Processing Units (GPUs) for those interested in machine learning models. All software on the system will go through a full risk analysis before installation.

Research Data Storage platform

Research Data Storage (RDS) is a Rutgers service offering research labs and university centers reliable data storage, with 34PB of usable space designed for both short- and long-term needs. Optimized for cost-effectiveness, it ensures faculty and principal investigators (PIs) reliable access to university research data. An on-premise solution that scales to handle large research datasets beyond conventional storage capacities, RDS features a three-site object storage system and two-site redundant storage configuration which provide high data resilience and availability. Designed to be user friendly for access and data management, RDS storage efficiency is optimized via standard protocol access, an active-active configuration, and end-user self-management.

RDS is a collaborative effort managed and maintained by OIT's Office of Advanced Research Computing (OARC) and Enterprise Infrastructure.

REDCap

Research Electronic Data Capture, known as REDCap, is a research-enabling web platform for building online databases and surveys and offers Rutgers researchers a secure environment for collecting and managing that survey data. The Office of Advanced Research Computing leads the support for the university-wide REDCap platform, which includes the following features:

- Streamlined process for rapid project creation with a vast array of tools for virtually any data collection strategy.
- Automated export procedures for seamless data downloads to Excel and statistical packages like SPSS, SAS, Stata, and R.
- Built-in project calendar, a scheduling module, ad hoc reporting tools, and more.
- Advanced features, such as branching logic, file uploads, and calculated fields.

Software

Software available to users of the Amarel system includes:

- Intel, GNU, and NVIDIA software engineering library and tool suites
- An Open OnDemand interface system to support graphical desktops and GUI-based or web-based applications (like Jupyter notebooks and RStudio)
- A FastX graphical desktop system (an alternative to Open OnDemand with unique application hosting capabilities)
- Common commercial software packages like MATLAB, Mathematica, Maple, SAS, SPSS, Stata are available with a proper software license provided by users
- A large collection of discipline-specific commercial packages like Amber, Gaussian, Q-Chem, MOE, and CLC Genomics Workbench

- An Apptainer container system for running Singularity or Docker images
- A variety of Conda-based environments and pre-configured data science software stacks
- A very large and constantly growing collection of open-source packages representing nearly all disciplines

Support and Education

OARC's Research Scientist team includes 16 experienced independent investigators with backgrounds in biochemistry, biomedical engineering, clinical informatics, chemistry, computer science, data science, image processing, computational genomics, geospatial information systems, mathematics, microbiology, molecular modeling, mechanical engineering, physics, and much more. This team provides multi-tier support for all users of the Amarel cluster ranging from assistance with installing software and running jobs to dedicated scientific support for funded and unfunded research projects. While several Research Scientist team members are dedicated to the CRDW and REDCap resources managed by OARC, they frequently cross boundaries between HPC, clinical, and REDCap projects, working as a team wherever their expertise is needed.

The Research Scientists also offer regular and customized education opportunities including but not limited to workshops, demonstrations, and guest speaker engagements at Rutgers and in the national community.

Events that are available or have been provided include:

Introduction to Linux from the Command Line
 Introduction to Research Computing
 Introduction to the Amarel Cluster
 Graphical / Interactive Sessions on the Amarel Cluster
 Computational Genomics: Variant Analysis
 Computational Genomics: RNA-Seq Analysis
 Managing Software and Data on the Amarel Cluster
 Intel FPGA Programming
 Linux Clusters Institute
 Molecular Dynamics and Simulation Engine Performance Tuning
 Introduction to Ab Initio and Density Functional Methods on Amarel
 Data Management Best Practices for Research
 Intel Tools and Libraries for Computational Chemistry
 Large-Scale Analytics Best Practices
 Introduction to Python
 Python for Big Data
 Cloud 101 for Research
 Cloud Best Practices for Research
 REDCap 101
 REDCap Intermediate Topics

REDCap Project Migration

Data in motion: utilizing the science DMZ and DTNs for optimal performance

Mastering data transfers on HPC systems

Supercharge your research: harnessing HPC resources

Data Centers

Rutgers has seven data centers geographically distributed totaling about 23,000 sq ft and a combined electrical input of 1.1 MW. OARC's anchor facility is located at the Hill Research Computing Data Center which provides 200 KW of power and attendant cooling. Most power is UPS-supported with backup diesel generators.

Location	UPS Power (kilowatts)	Cooling (tons)
Piscataway (IBM Modular Data Center, Hill Center)	600 (200 non-UPS)	~225
Camden	35	10
Newark (Engelhard Hall)	52	30

Campus and Regional Networks

Production Network

Rutgers University's production data network, RUNet, is a purpose-built network of nearly 100 routers and 2,500 switches servicing constituents of 8 campuses across three metropolitan areas and several dozen remote locations throughout New Jersey. The network interfaces with service providers at two of its main Points of Presence (POPs) in Philadelphia, PA and Newark, NJ. RUNet provides access to 30 Gbps of Internet transit, 10 Gbps of Internet2 IP transit, 10 Gbps of Internet2 Advanced Layer 2 Services, 20 Gbps of private Internet IP Exchange services and several private network interfaces to select content providers.

Rutgers is also an anchor member of the NJEdge consortium, a collaboration network of NJ Higher Education.

Research Network

Funded in part by the National Science Foundation (Grant #OAC-1659232) and jointly supported by OARC and the telecom division of the Office Information Technology, the Rutgers Research Network (originally referred to by the grant project name, CICNet (NSF Award 1659232) is a 100 Gbps SDN network offering Rutgers investigators advanced compute and file access services while also serving as a testbed for new technologies and high-speed connections to local, regional, and national services in academic and commercial domains.

Regional Resource Collaborations

Rutgers took a leadership role in establishing the **Ecosystem for Research Networking**

(**ERN**, formerly known as the Eastern Research Network). Through a partnership of educational institutions, research facilities, regional network providers, and Internet2, the ERN is committed to providing layered and transparent access to shared data and computing facilities for research projects located at member sites. As the ERN developed, research communities across the region have been given access to a broad range of services and resources that may not be available on any one campus alone. Academic institutions, research and educational networks and corporate participants are all working together to demonstrate the value of the ERN. Penn State University, University of Massachusetts Amherst, Syracuse University, University of Maine, New Jersey Institute of Technology, Massachusetts Green High Performance Computing Center, OSHEAN, NJEdge.Net, Kinber, Google, Lenovo and SchedMD are among the expanding list of ERN member organizations.

The **New Jersey Big Data Alliance (NJBDA)** was established by the Rutgers Office of Research and Economic Development and the Rutgers Discovery Informatics Institute to catalyze the collaboration among New Jersey government, academia and industry that will enable all parties to address the significant and immediate challenges posed by the proliferation of data sources and the resultant deluge of digital data in a strategic and coordinated manner. This unprecedented alliance brings together universities and colleges from across the state, and has the overarching goals of identifying common challenges and areas of synergy, developing joint programs, and ultimately nucleating an effective alliance that will increase our research competitiveness and drive economic development in New Jersey. Founding members include Kean University, Montclair State University, N.J. Institute of Technology, Rowan University, Richard Stockton College, **Rutgers University** and Stevens Institute of Technology. The Alliance also works closely with the Office of Assemblyman Upendra Chivukula, the New Jersey Office of Information Technology - Enterprise Data Services, the New Jersey Economic Development Authority, and NJEdge.net, New Jersey's Higher Education and Research Network.

Other Resources

The **Rutgers Office of Information Technology (OIT)** develops, implements and supports shared services for the entire Rutgers Community. It has been recognized for years by the New Jersey Office of Higher Education as innovators in advanced technologies and often plays a leadership role in state-wide initiatives. OIT employs hundreds of staff and students supporting all areas of administration, on-line instruction, research computing services all connected by a network infrastructure which is one of the largest in the state of New Jersey. While each campus and school maintains their own IT organizations, distributed units collaborate with OIT on shared infrastructure, services and technological know-how.

As Senior Vice President for Information Technology at Rutgers and Chief Information Officer, **Michele L. Norin** is Rutgers' institutional leader for technology. Her chief responsibility is to provide leadership in the strategic adoption and use of information technology in support of the University's vision for excellence in research, teaching, outreach, and lifelong learning. Michele

will support this project by ensuring the resources needed to reach the objectives are available, including but not limited to human, financial, and equipment resources.

The **Office of Advanced Research Computing (OARC)** is a division of OIT, within **Enterprise Applications and Research Computing** led by Vice President **Ellen Law**. As Vice President, Ellen provides executive leadership and oversight for university-wide Enterprise Applications and Advanced Research Computing at Rutgers, partnering with academic, research and administrative units to further the university's mission of excellence in research, teaching and outreach. The enterprise software that she oversees assists with ensuring the work of the university is executed in a strategic, secure, and cost-effective manner. OARC is the culmination of several years' effort to coalesce research computing activities and assets into a program with a coordinated strategy to provide Rutgers investigators with a state-of-the-art environment designed to address the most demanding problems by seamlessly integrating local, regional, national and cloud computing resources. Technological evolution and sustainability are a cornerstone of this effort.

As Associate Vice President for Research Computing, **Juan Perin** leads the strategy, operations, and growth of advanced research computing at Rutgers, partnering with faculty, researchers, and administrative units to advance the university's mission of excellence in research, teaching, and outreach. He oversees the high-performance computing, data, and research infrastructure that enables Rutgers researchers to pursue ambitious, computationally intensive work in a secure, scalable, and cost-effective manner. Juan brings extensive experience in research computing and IT leadership, having held a range of technical and leadership roles that inform a hands-on, researcher-focused approach to building the systems that power discovery at Rutgers.

OARC's Director of Research Support, **Vladyslav Kholodovych, PhD**, leads the Research Education and Facilitation (REF) team, as well as the CRDW and REDCap teams at OARC. Dr. Kholodovych's primary responsibility is to ensure the seamless integration of high-performance computing resources into ongoing research endeavors while prioritizing the protection of sensitive data and strict regulatory compliance. He coordinates the utilization of OARC's cyberinfrastructure resources, ensuring their use aligns with the regulations governing a protected data environment. This includes implementing measures to safeguard sensitive information and facilitating the necessary compliance protocols. Beyond core HPC services, he also oversees OARC's implementation of other critical resources, including the Clinical and Research Data Warehouse (CRDW), REDCap for secure data collection and management, and the Research Data Storage (RDS) system, ensuring comprehensive support for the university's research community.

In addition to this role within OARC, Dr. Kholodovych holds two adjunct positions: Associate Professor in the Department of Medicinal Chemistry at the Ernest Mario School of Pharmacy, and Assistant Professor in the Department of Pharmacology at Robert Wood Johnson Medical School. He is actively involved in teaching graduate-level courses on Structural Bioinformatics and Drug Design and regularly serves as a scientific advisor on graduate and PhD committees.

As a recipient of the Humboldt Fellowship, Dr. Kholodovych contributed to the development of the Online Chemical Modeling Environment (OCHEM), a web-based platform for data storage, model development, and the publication of chemical information.

OARC's Security and Research Architect, **Eric Marshall**, has built complex systems for a quarter century for finance giants like Oppenheimer and Bloomberg and research platforms like NOAA's N-Wave backbone and Geophysical Fluid Dynamics Laboratory's Top500 supercomputers. The former leader of the Advanced Computing Infrastructure group within the Office for Advanced Research Computing, Eric has overseen the architecture, maintenance, and daily operations of a research-intensive compute, storage, and data network. Among many other certifications, Eric is a certified AWS Cloud Practitioner and an ISC2 Certified Cloud Security Professional.

OARC's Director of Strategic Engagement and Research Partnerships, **Kristin Lepping**, MLIS, leads strategic initiatives, oversees OARC's grant strategy, and guide programs in education, training, student engagement, outreach, and communications. Kristin acts as OARC's bridge to funding agencies, providing the team with research compliance oversight and guidance, while also working with internal leadership to develop new projects and their business models, writing policy and ensuring top of the line customer service. Her emphasis at OARC and in the research computing community is on the people who make discovery possible through technology—the experts who enable and engage directly in research, support scholars, and grow the field.

Kristin's national outreach has included contributions to the Practice and Experience in Advanced Research Computing (PEARC) Technical Program and serving as Co-Awards Chair; serving on the Coalition for Academic Scientific Computation (CASC) Communications and Program committees; participating in the EDUCAUSE Research Computing and Data Community Group Steering Committee; and serving as Administrative Director of the New Jersey Big Data Alliance. Her efforts are focused on expanding opportunity, strengthening community, and ensuring that the research computing field remains inclusive, visible, and impactful.