

The Center for Research Informatics (CRI) is dedicated to providing resources to University of Chicago faculty to enable biological research. The CRI is comprised of four service lines: (1) data warehousing, (2) scientific computing and research storage, machine-learning, and visualization, (3) application and platform development, and (4) bioinformatics.

Research computing infrastructure and resources for the Center for Research informatics (CRI) are primarily located in the **Kenwood Data Center** which is in Hyde Park on the University of Chicago campus. The Data Center is designed and tested to withstand extended power outages without system or service interruption. CRI servers are physically secured in locked racks in a facility fitted with electronic entry and alarm systems. The data center is divided in two sections, each designed for different use profiles: POD-A (2,500 square feet) and POD-B (2,100 square feet). POD-A is designed to house mission-critical workloads and meets the Uptime Institute's Tier 2 rating. POD-B is designed for compute-heavy High-Performance Compute (HPC) with a power draw of up to 25kW per cabinet. The Kenwood data center is managed by University of Chicago IT Services and monitored by staff 24/7. The datacenter is equipped to house systems that may fall under certain federal guidelines, including Health Insurance Portability and Accountability Act (HIPAA) and the Federal Information Security Management Act (FISMA). The **1155 Data Center** is also located in Hyde Park on the University of Chicago campus, which is 4,144 square feet and managed by University of Chicago IT Services. The CRI utilizes this facility to house the tape library that is used for data backups.

Compute Resources: The CRI provides comprehensive computing infrastructure resources that include a High performance Computing (HPC) Cluster ("Randi") which contains 211 compute nodes, equipped with: 2 intel ice lake processor with 16 cores per processor and 3 memory tiers (128 GB, 512 GB, and 1280 GB), 100+ bioinformatics software applications, commercial and open-source compilers for software development, integration with Galaxy, 350 TB scratch space, and a web-based portal for biomedical analysis. The cluster, Randi, also contains 5 specialized GPU nodes that are equipped with 2 AMD Milan processors with 16 cores per processor and 8 Nvidia A100 GPUs and 1 SXM GPU node equipped with 2 AMD Milan processors with 16 cores per processor, 8 NVidia A100 GPUs and NVLink technology for high-speed interconnectivity between GPUs.

The CRI, in partnership with BSD Information Services manages a **Virtual Server Infrastructure** with around 450 active virtual servers on VMWare VSphere (Dell VxRail hardware platform).

The CRI provides 6.5 petabytes of Secure, HIPAA-compliant **File Storage** (IBM ESS storage system using GPFS) with available space to be provisioned for group shares assigned to BSD principal investigators. To ensure that access is only granted to authorized staff, all changes to group share access are verified with the principal investigator or designee. The storage systems are backed up to a SpectraLogic T950 tape library using IBM Storage Protect. The tape library is housed in a datacenter in a different location on campus to ensure separation from the original data. SpectraLogic Storcycle is used for **Data Archiving** functionality, where research data is archived to Microsoft Azure Blob storage *and* on tape in the SpectraLogic T950 tape library.

These resources collectively provide the necessary power and versatility to support a wide range of computational tasks, including bioinformatics analyses, software development, and deep learning applications.