FACILITIES, EXISTING EQUIPMENT, AND OTHER RESOURCES

The University of Chicago is a private non-profit institution located on the ethnically-diverse South Side of Chicago that has been a center of advanced learning and research since its inception in 1892. The University of Chicago is comprised of four graduate Divisions (Biological Sciences, Physical Sciences, Social Sciences, and Humanities), six professional schools (Chicago Booth School of Business, Divinity School, Harris School of Public Policy Studies, Law School, Pritzker School of Medicine, and School of Social Service Administration), the Graham School of General Studies, and the undergraduate College. The University has a unique history of organizing around research questions that cross disciplines rather than operating within disciplinary boundaries. The extent to which this strategy reflects University of Chicago is illustrated by its numerous interdisciplinary Committees, Centers and Institutes (described below). The University of Chicago maintains its commitment to scholarship, teaching, and research through its more than 2100 faculty members and a student population of approximately 15,600 with nearly 2/3 engaged in advanced research and professional study. Through the years, 86 Nobel Laureates (8 are current faculty), 44 members of the National Academy of Sciences, 169 members of the American Academy of Arts and Sciences, and 14 recipients of the National Medal of Science have been associated with the University as students, teachers or research investigators. The University of Chicago is ranked among the world's top universities by a number of criteria, including the amount of federal research funding received (despite a size much smaller than many of its academic peers). This spirit of discovery, innovation and public service provides a robust foundation for success.

South Shore Senior Health Center (SSSC): The SSSC is a 6800 sq. ft. university-owned geriatric facility 5 miles south of DCAM, with doorstep free parking in the heart of Chicago's South Side, specifically in the South Shore district. There are 13 patient exam rooms equipped with an examination bed, ophthalmoscope and otoscope, in addition to an on-site phlebotomist and capability for ECG. There is a large conference room for team meetings and support groups. The Memory Center team meets on Mondays at this site.

Center for Care and Discovery: Completed in 2013, the CCD is a ten-story adult hospital focused on cancer, advanced surgery, high tech imaging, the neurosciences and gastrointestinal procedures. The building is 1,200,000 GSF with floor plates of 102,000 GSF. The CCD includes 240 private patient rooms, 28 operating rooms (21 initially); an imaging department with 3 CT's, 2 MRI's, 1 fluoroscopy room, 2 general radiology rooms, 7 interventional radiology rooms; and a gastroenterology procedures suite with 11 GI procedure rooms, 2 fluoroscopy rooms and 2 bronchoscopy rooms. The CCD includes an inpatient kitchen, cafeteria, 7th floor sky lobby meeting rooms, ground floor retail space and clinical support services. Two floors are "shelled space" for future expansion of services. The building is centrally located between existing clinical facilities (DCAM, Comer) and new research facilities (KCBD, GCIS, Knapp). The facility is connected to both DCAM and Comer via above and below ground connections. There are procedure rooms, 2 fluoroscopy rooms and 2 bronchoscopy rooms. Neurology and neurosurgery inpatients as well as a Neuro- ICU are contiguously located on one floor of the CCD.

COMPUTATIONAL FACILITIES

The principal investigators have access to extensive computational facilities available at the University of Chicago to carry out the tasks described.

Access to Clinical Data for AI-enabled Analytics: The ZeD lab (overseen by Professor Chattopadhyay) is housed within the Department of Medicine at the University of Chicago, and has access to the full range of high end computing resources offered by the University of Chicago. In addition, Prof. Chattppadyay's laboratory has access to the HIPAA compliant clinical data warehouse maintained by the Biological Sciences Division as detailed below:

The Clinical Research Data Warehouse: (CRDW) within the Biomedical Sciences Division of the University of Chicago is one of the deepest, richest, and most research-ready data repositories of its kind. Containing more than a decade of University of Chicago medical data, it seamlessly brings together multiple internal and external data sources to provide researchers with access to more than 12 million encounters for 2.3 million patients. The associated diagnoses, labs, medications, and procedures number in the tens of millions each. The CRDW is run on IBM Netezza Pure Data System for Analytics servers, a patented Asymmetric Massively Parallel Processing architecture designed to deliver exceptional query performance and modular scalability on highly complex mixed workloads.

TABLE 1
University of Chicago Research Computing Center Capabilities Summary

Cluster	Partition	Compute cores (CPUs)	Memory	Other configuration details
midway1	westmere	12 x Intel X5675 3.07 GHz	24 GB	
	sandyb	16 x Intel E5-2670 2.6GHz	32 GB	
	bigmem	16 x Intel E5-2670 2.6GHz	256 GB	
		32 x Intel E7-8837 2.67GHz	1 TB	
	gpu	16 x Intel E5-2670 2.6GHz	32 GB	2 x Nvidia M2090 or K20 GPU
		20 x Intel E5-2680v2 2.8GHz	64 GB	2 x Nvidia K40 GPU
	mic	16 x Intel E5-2670 2.6GHz	32 GB	2 x Intel Xeon Phi 5100 coprocessor
	amd	64 x AMD Opteron 6386 SE	256 GB	
	ivyb	20 x Intel E5-2680v2 2.8GHz	64 GB	
midway2	broadwl	28 x Intel E5-2680v4 2.4GHz	64 GB	
	bigmem2	28 x Intel E5-2680v4 @ 2.4 GHz	512 GB	
	gpu2	28 x Intel E5-2680v4 @ 2.4 GHz	64 GB	4 x Nvidia K80 GPU

In order to meet the acute need for data related to COVID-19, the CRDW team has constructed three data marts (de-identified, limited, and identified) to provide the most commonly requested data elements for this patient population. The initial instance of the COVID-19 data mart includes de-identified structured data on patient demographics, encounters, diagnoses, labs, medications, flow sheets, and procedures. Additional data will be added based on resource availability and urgency.

Cohort Discovery Tool: The purpose of this tool (SEE Cohorts) is to provide a secure web-based tool for the initial exploration of de-identified data. It allows researchers to search available data, build a cohort of patients, and view actual de-identified data within the interface. The data in SEE Cohorts is refreshed weekly.

Research Computing Center: The University of Chicago Research Computing Center (RCC) provides high-end research computing resources to researchers at the University of Chicago, which include high-performance computing and visualization resources; high-capacity storage and backup; software; high-speed networking; and hosted data sets. Resources are centrally managed by RCC staff who ensure the accessibility, reliability, and security of the compute and storage systems. A high-throughput network connects the Midway Compute Cluster to the UChicago campus network and the public internet through a number of high-bandwidth uplinks. To support data-driven research RCC hosts a number of large datasets to be accessed within the RCC compute environment.

RCC maintains three pools of servers for distributed high-performance computing. Ideal for tightly coupled parallel calculations, tightly-coupled nodes are linked by a fully non-blocking FDR-10 Infiniband interconnect. Loosely-coupled nodes are similar to the tightly-coupled nodes, but are connected with GigE rather than Infiniband and are best suited for high-throughput jobs. Finally, shared memory nodes contain much larger main memories (up to 1 TB) and are ideal for memory-bound computations. The types of CPU architectures RCC maintains are tabulated in Table 1.

RCC also maintains a number of specialty nodes:

- *Large shared memory nodes* - up to 1 TB of memory per node with either 16 or 32 Intel CPU cores. Midway is always expanding, but at time of writing RCC contains a total of 13,500 cores across 792 nodes, and 1.5 PB of storage.
- *Hadoop:* Originally developed at Google, Hadoop is a framework for large-scale data processing.
- *GPU Computing:* Scientific computing on graphics cards can unlock even greater amounts of parallelism from code. RCC GPU nodes each include two Nvidia Tesla-class accelerator cards and are integrated in the Infiniband network. RCC currently provides access to Fermi-generation M2090 GPU devices and Kepler-generation K20 and K40 devices.
- *Xeon Phi:* The Many Integrated-Core architecture (MIC) is Intel's newest approach to manycore computing. Researchers can experiment with these accelerators by using MIC nodes, each of which have two Xeon Phi cards, and are integrated into the Infiniband network.

Persistent and High-Capacity Storage. Storage is accessible from all compute nodes on Midway1 and Midway2 as well as outside of the RCC compute environment through various mechanisms, such as mounting directories as network drives on your personal computer or accessing data as a Globus Online endpoint (at the time of this writing, Globus Online is supported on Midway1). RCC takes snapshots of all home directories (users' private storage space) at regular intervals so that if any data is lost or corrupted, it can easily be recovered. RCC maintains GPFS Filesystem Snapshots for quick and easy data recovery. In the event of catastrophic storage failure, archival tape backups can be used to recover data from persistent storage locations on Midway. Automated snapshots of the home and project directories are available in case of accidental file deletion or other problems. Currently snapshots are available for these time periods: 1) 7 daily snapshots, 2) 4 weekly snapshots.

Tape Backups. Backups are performed on a nightly basis to a tape machine located in a different data center than the main storage system. These backups are meant to safeguard against events such as hardware failure or disasters that could result in the complete loss of RCC's primary data center.

Data Sharing. All data in RCC's storage environment is accessible through a wide range of tools and protocols. Because RCC provides centralized infrastructure, all resources are accessible by multiple users simultaneously, which makes RCC's storage system ideal for sharing data among your research group members. Additionally, data access and restriction levels can be put in place on an extremely granular level.

Data Security & Management. The HIPAA compliant security of the Research Computing Center's storage infrastructure, protected by two-factor authentication, gives users peace of mind that their data is stored, managed, and protected by HPC professionals. Midway's file management system allows researchers to control access to their data. RCC has the ability to develop data access portals for different labs and groups.