March 2016 Progress Report

Executive Summary

Cyberinfrastructure is broadly defined as the human and technological support framework for advanced data acquisition, data storage, data management, data integration, data mining, data visualization, data curation and other computing and information processing services within the research environment.

Research Computing has been growing to meet the needs of researchers on campus with a number of improvements, new services, and a new direction. Highlights of 2015 include the following:

- More than one petabyte of newly installed General Purpose Research Storage has been installed on campus to address the urgent need for research data storage.
- The High Performance Computing (HPC) cluster has been upgraded with more capacity and an updated architecture, and user training has been expanded.
- MU faculty has helped guide the introduction of a number of grant-friendly services to help researchers gain computing capacity without having to worry about managing the underlying infrastructure.
- A next generation 100-Gigabit Software Defined Networking (SDN) high-speed network has been installed to address the future needs of researchers and their need to access off-campus resources and remote collaboration.

Through all these changes, and with the guidance of MU's Cyberinfrastructure Council, the Division of IT is striving to better support MU researchers with their computational needs.

Background

CI Council members provide input and guidance on the cyberinfrastructure needed to support the wide variety of research and discovery activities across the University of Missouri. Their input is critical for planning strategic investments and ongoing support for cyberinfrastructure, as well as the services and resources provided by the Division of IT's Research Computing Support Services group.

The CI Council is comprised of faculty and staff representing all MU Schools and Colleges and meets monthly.

Progress Towards Goals

MU's Campus Cyberinfrastructure Plan

Shortly after forming in 2013, MU's CI Council adopted the campus' first plan and set priorities for research cyberinfrastructure. That plan laid out principles and expectations. (doit.missouri.edu/research/ci). As progress continued, the group determined a need for a more specific and measurable plan. Following a large-group discussion, smaller groups were formed to address these particular facets of the plan:

- Bioinformatics & Genomics
- Digital Humanities
- Geospatial Sciences
- Imaging and Visualization

These groups met during the summer and fall of 2015 to determine the appropriate direction, establish priorities, and to craft a document. Discussions continue in 2016.

CI Plan Update – Discussions and Presentations

Date	Facet
November 2015	Geospacial Sciences
December 2015	Digital Humanities
January 2016	Bioinformatics and Genomics
February 2016	Imaging and Visualization

Each group has prepared an update to their facet of the plan to share with the campus on CI Day 2016. The Council will then move toward securing administrative buy-in to the plan and the associated funding. The latest plan is available at ***doit.missouri.edu/ci/ci-plan***.

Enhancing Communications

Researcher Presentations

Discussions with MU researchers regarding their research and the related needs for cyberinfrastructure are a regular agenda item at the monthly meetings. Each researcher makes a brief presentation designed to stimulate discussion. To maximize the consistency and value across presenters, each is asked to dedicate one or more slides to these topics:

1. Provide a brief overview of your research interest, why it's innovative, the future trajectory, and relevant funding.
2. Provide information on research collaborators and their resources relevant to your research.
3. Describe Intellectual Property (IP) issues.
4. Describe you CI 'pain points' and how they affect your research.
5. If money were no object, what CI environment and resources would you have?

The presentations provided a catalyst for an engaging discussion that provided insight about the researchers work and their Cyberinfrastructure needs.

MU CI Council – Researcher Presentations & Discussions

Date	Researcher	Topic or Title
December 2014	Sean Coggins	Dr. Coggins' NSF-funded research on how groups use IT for communications
January 2015	Chris Fulcher	Community Commons communitycommons.org/
March 2015	Jeff Rouder	Open Data Exchange "The What, Why, and How of Born Open Data"
April 2015	Jim Bottum, Clemson University	Follow-up conversation to 3-3-15 CI Day
May 2015	Timothy Haithcoat	Geospatial Sciences
June 2015	Abu Mosa	Computing in Clinical and Translational Research
August 2015	Kathleen Slane	More Meaningful Dissemination of Archaeology Research
September 2015	Diane Oerly	Briefing on Research Data Alliance
October 2015	Nathan Bivens	Briefing on MU's DNA Core

DoIT Website

- Dr. Timothy Middelkoop, Director of Research Computing Support Services, continues to use his **Research Computing News** blog to keep readers up-to-date with the latest research activities and events at the University of Missouri.
- The Division of IT continues to maintain a web presence for the **CI Council** and works to keep the MU CI Resources Guide up-to-date.

CI Day 2015

The Cyberinfrastructure Council held its second CI Day on Tuesday, March 3, 2015. CI Day fosters collaboration, networking, and collective problem solving.

Attendees were able to learn about advanced computing technologies across a wide range of disciplines. The program focused on data, with topic tracks spanning a variety of disciplines and resources. Sessions included:

- Security: Threats to Intellectual Property
- MU Cyberinfrastructure Resources to Enable Data-intensive Research and Education
- DAMS (Digital Asset Management Systems)
- Open Science Framework
- Hathi Trust Research Center
- New Certificate and Degree Program Offerings

CI Day 2016

Plans are underway for MU's third CI Day to be held April 13, 2016. Henry Neeman of the University of Oklahoma and Seth Colaner from Tom's Hardware have agreed to be the keynote speakers. They will meet with CI Council members and others while they are on campus.

Increasing Funding for CI

In 2015, campus administration allocated $500,000 in one-time funds to increase research data storage capacity and $750,000 in recurring funds to enhance technology and staffing for research support. No additional funds have been allocated for Cyberinfrastructure in 2016; however, this continues as a main focus of discussion by administration.

Responding to MU Researcher's Most Critical Needs

Research data storage and more High Performance Computing (HPC) capacity emerged as the top priorities for MU campus investments. The following services were launched in 2015:

General Purpose Research Storage (GPRS)

In 2015, the Division of IT purchased more than one petabyte of storage and, for the first time, offered general purpose research storage. The GPRS is a sustainable and flexible research data storage environment offered at minimal cost. The CI Council was instrumental in setting the policy for this new resource.

The Research Computing staff has interconnected and will manage the data storage and actively recruit additional investments by researchers and other units on campus.

The GPRS investment and cost structure make it possible to budget for a projects' storage costs during the proposal development process, and include these costs in the grant budget.

Storage Offerings Include:

- **Individual Storage**: Individual researchers are given 10 GB of private storage for private use at no charge. Additional storage is allocated in 256 GB increments at the cost of $10/TB/month.

- **Project Storage**: Storage associated with specific groups of researchers or instruments are allocated on a per project basis. Each project is provided with 10 GB of storage at no charge. Additional storage will be allocated in 256 GB increments at the cost of $10/TB/month.

- **Special Project Storage**: Fifty TB of storage have been designated for special cases where the storage will be provided at a reduced cost or at no charge. These projects are expected to be short exploratory projects or special projects evaluated on a case-by-case basis considering the need (size and duration). Special Projects might include promising exploratory research for short periods of time, exemplary research that has little chance of external funding or needs proven results to secure funding, and other special-use cases as determined by MU's CI Council.

- **Large Storage Investment**: Researchers with larger storage needs can invest in half or multiple entire nodes, around 100 TB per node, for five years of service. Investors receive dedicated access to one (half-node) or two (full-node) 10-Gigabit Ethernet (GbE) network storage ports. Nodes will be taken out of service after five years, but as part of a future storage offering, may be placed in a degraded storage system.

General Purpose Research Storage | *General Purpose Research Storage Policy*

High Performance Computing

High Performance Computing (HPC) is a system-wide service that provides the advanced computational and data-driven capabilities needed for innovative and collaborative research activities at the University of Missouri. The HPC environment includes a state-of-the art shared-resource cluster, an experimental cluster, and a number of grant-friendly investor services. A teaching and learning cluster for students is also under development.

Acquisitions and Improvements to Hardware

Shared-Resource Cluster (Lewis3):

The current cluster has 960 Haswell cores (with between 128 and 512 GB of RAM), connected with 10-Gigabit Ethernet, 40-Gigabit Infiniband, and connected to two PB of storage with 100 TB of high-speed scratch. The cluster runs CentOS 7 with the SLURM job scheduler.

The cluster supports more than 200 scientific software packages from a large number of disciplines. We will install most open source software packages and libraries on an as-needed basis and can host licensed software as well. We have licenses for SAS, Matlab, Gaussian, Intel compilers and Parallel Studio, as well as others.

Experimental Cluster (NSF MRI):

The University of Missouri hosts an experimental cluster (NSF MRI Award #1429294) used to provide unique capabilities to researchers working on and with high performance research computing. This cluster is designed and configured to be highly flexible to allow specialized and dedicated environments for researchers. This means researchers are able to conduct research at scales not available in lab-scale systems and in specialized configurations that are coordinated by an allocation board of faculty, staff, and students. Configurations can test scalability and deploy large specialized environments such as Openstack, Hadoop, Spark, and others, along with a traditional HPC configuration.

The cluster is in the first phase of deployment and currently has a total of 956 cores on 46 nodes with dual socket 20 and 24 Haswell cores (E5-2670 v3 2.30GHz and E5-2670 v3 2.30GHz) with 128 GB and 256 GB of DDR4 RAM, with two high memory nodes with 512 GB of RAM. 16 nodes are accelerator nodes with 16 Intel Phi cards, 8 K20m, and 2 K40 GPUs. Nodes can be configured between 1T and 12T of local storage (total of 6.5T of RAM, 86 TB of storage, and 2.6 TB of SSD) with access to 1 PB of external storage.

The cluster supports multiple networking fabrics to facilitate experimentation (multiple 1- Gigabit management networks, 10-Gigabit Ethernet with SDN capability, and 40-Gigabit QDR Infiniband fabrics) and is connected to the campus 100-Gigabit Internet2 AL2S connection.

The cluster is managed by the Division of IT Research Computing Support Services group.

HPC Investor Machines

The HPC Investor machines are Dell R630 1U compute nodes with 128 GB of DDR4 RAM and 24 Haswell cores (2x E5-2670v3, 2.3Ghz) with 1 TB of local storage with 10-GB Ethernet and 40-GB QDR Infiniband networking.

Investors purchase the node hardware; however, the management of rack space, power, and cooling services along with hardware, operating system, and applications is provided by Research Computing Support Services at no cost for five years. Other configurations are available.

Training

Research Computing Support Services provides ongoing training for Lewis 3 each week. This training series covers the basics of the new hardware, how to use the new scheduler, how to use secure shell key-based authentication, and any other questions researchers may have.

Last October, Research Computing Support Services hosted an all-day training session that featured presentations and hands-on exercises on the available programming models and best optimization practices for the Intel Xeon Phi coprocessors, and on the usage of the Intel software development and diagnostic tools.

Proposed Master's Degree in Data Science and Analytics

MU has proposed a new Masters Degree in Data Science and Analytics to be offered through collaboration among MU schools and colleges of Engineering, Journalism, Education, and the MU Informatics Institute (and eventually others). The degree proposal has been approved by the University of Missouri's Board of Curators, and is nearing the final step in the approval process. It is on the agenda for the April, 2016 meeting of Missouri's Coordinating Board for Higher Education (CHBE).

CI Council members represent each of MU's Schools and Colleges. For questions or concerns about MU's CI Council, email: ***MUCIDAY@missouri.edu***.