

University of Missouri – Cyberinfrastructure (CI) Progress Report for 2016 and 2017
Appendix A – Updates by Semester provided by Dr. Timothy Middelkoop
Online at: <https://doit.missouri.edu/research/research-computing-news/>

Research Computing Support Services Winter 2016 Update

The past year has seen a lot of changes for research computing on campus and a lot are in the works. The beginning of the year saw the commissioning of more than one petabyte of General Purpose Research Storage (GPRS) for researchers on campus to have an affordable way to store large amounts of data (currently at \$10/TB/month). The MU Cyberinfrastructure Council was instrumental in the development of the policy for the storage system and will help develop new services and policies in the future. The Lewis HPC upgrade started with the commissioning of 32 new nodes in conjunction with a new experimental cluster and 32 additional nodes sponsored by the National Science Foundation. The combined system has more than 2,000 cores and is growing. Because the two systems share infrastructure, the Lewis3 system can take advantage of cycles not used by grant activities.

In addition to the GPRS service, we launched our HPC Investor service, which allows researchers to invest in the cluster in a grant-friendly way and get priority access to those resources. This partnership allows researchers to obtain computational capacity without having to manage their own systems. For more information see <https://doit.missouri.edu/services/research/high-performance-computing>.

The year 2015 also saw the transition of the stewardship of the research computing infrastructure from Research Support Computing (RSC) led by Gordon Springer to Research Computing Support Services (RCSS) led by myself, Timothy Middelkoop. Gordon Springer retired after many years of dedicated service to the campus community; however, he is still active in the community as an Emeritus faculty in the Computer Science department. The new team and I look forward to building upon the work that Gordon and his team did.

As we look forward into the new year we will be upgrading a lot of our infrastructure. Of note we will be:

- Upgrading the HPC storage system
- Decommissioning the old Lewis cluster (Lewis2) this summer
- Upgrading and migrating services connected with the Lewis2 cluster

This will have an impact on a lot of existing services, so we will attempt to contact all those involved. Researchers that have invested in the past with RSC, or have active services, are encouraged to contact us directly to develop a migration plan.

We are also building new services to serve researchers on campus in a sustainable way. Of note we are:

- Building an OpenStack environment to provide virtual machines for researchers to be used in their research
- Building a Secure Research Compute cluster for researchers that need a secure compute environment for DCL3 and DCL4 data (<http://infosec.missouri.edu/classification/dcs.html>)
- Building a student cluster to provide computational resources for teaching
- Upgrading the campus Research Network (RNet) infrastructure that provides a Science DMZ for researchers on campus.

As a part of the upgrade to Lewis3, we have been providing training sessions for new and existing users to get the most out of the cluster. Last fall we also brought Intel to campus for a course on parallel computation. We plan on expanding the training opportunities for researchers on campus in the upcoming year. We will also be holding periodic user forums where we can talk with researchers about the cluster and to help improve our services. Watch this space for updates.

RCSS Summer 2016 Update

Research Computing Support Services continues to grow in terms of people, equipment, and infrastructure. First of all, I would like to welcome our latest team member, Jacob Gotberg. Jacob will be joining us in July as a Cyberinfrastructure Engineer and will be equally supporting research computing on campus and the College of Engineering. Jacob joins us from Sandia National Laboratories where he worked in high- performance computing (HPC).

I would also like to announce the upgrade and an increase in capacity to the Lewis cluster. We will be adding 16 new nodes to the community compute pool to meet the increased computing demands of users and our HPC investors are adding at least eight additional investor nodes to the cluster. The decommissioning of the legacy Lewis2 cluster and associated storage will allow us to upgrade the provisioning system and integrate some of the older capacity (34 nodes/616 cores) into the cluster as well.

Now would be an excellent time to become an HPC investor as our large purchase has allowed us get nodes at a very competitive price. With the HPC investor program researchers purchase an HPC node and we provide, at no-charge, the space, power, and cooling. We also manage the hardware, operating system, scientific software, and security required to run the nodes. Investors get HPC group storage of 3TB at no charge as well. Please contact me directly if you are interested in becoming an HPC investor. We have also been busy building a teaching cluster (10 nodes) for students in courses to use for their scientific computational needs. This environment will be similar to the Lewis cluster, but tailored for instructional needs.

Over the past six months we have also been working on a Bio Compute cluster specifically targeted at Bioinformatics and Genomics workloads and workflows. Please contact us if you are interested in becoming an investor as well.

Since the last update we have had a number of disruptive storage events and we finally have a handle on the root cause. The issue was around how the different storage pools interact on the cluster (think of it as different disk drives in your workstation) and have now mitigated the problem. We would like to thank the community for both being very understanding and for helping us debug the issues. Without your flexibility and help it would have been a much more difficult series of events.

This is a good time to remind everyone that even though most of our systems are very resilient to component failure we do not backup users' data. We employ the philosophy of "Research Grade" where we try to maximize compute, storage, and researcher productivity, and this is one of the tradeoffs we make.

The decommissioning of the original Lewis infrastructure will occur on July 1, and allow us to free up five racks of compute and storage to make room for additional cooling in the datacenter. As a part of this the support@rnet.missouri.edu email address will no longer be active and users should use rcss-support@missouri.edu.

During the transition, we will be normalizing group storage and group management so we can automate many of our processes to serve researchers better and faster. We will be contacting individual groups to migrate their storage and groups to the new system. Over the next few weeks we will also be testing the new environment (hardware, software, and infrastructure) and inviting researchers to help us shake out the bugs. If all goes well, we will have an extended outage the first week of August to make the final changes to the environment.

The Research Network (RNet) will see a much needed upgrade in FY17. Even though we are still in the early planning phases, there are two important changes that will have an impact on users of RNet. The first is that almost all Gigabit Ethernet RNet ports will utilize Tigernet to reach RNet, and starting in FY18, all connections will be charged the standard port fee. As the Research Network is upgraded during FY17, all RNet port fees will be waived as connections are migrated. The second is that we will be rolling out a new RNet architecture to better secure the research network where it needs to be while keeping it open where it needs to be. As we upgrade RNet, we will be working with individual researchers to find the service that best fits their needs. For researchers with specialized networking needs (networking

research, SDN, 10Gigabit), we will continue to partner with them to find the best dedicated hardware solution available. Continue to watch this space for more information about the RNet upgrade. We are working hard on many other projects, so keep watching this space for future announcements about secure compute, bio compute, science gateways, and storage upgrades. And, as always, don't hesitate to reach out to me or one of the others on the team.

— Jun. 15, 2016

RCSS Fall 2016 Update

Research Computing Support Services (RCSS) continues to evolve to meet the needs of researchers on campus. Here are some of our latest activities:

HPC Environment Upgrade

This summer, we advanced the campus research computing, storage, and networking infrastructure by upgrading and expanding two existing clusters, building four new specialized clusters, and decommissioning old equipment and infrastructure. This involved:

- Decommissioning aging compute, storage, and support infrastructure to improve security, power usage, maintenance costs, space, and cooling.
- Implementing the next generation provisioning and configuration system to facilitate growth and investment.
- Investing in additional HPC capacity to serve the entire research community across campus and UM System.
- Increasing the amount of investor-purchased (colleges, labs, cores, and researchers) computing capacity. This included a sizeable purchase for the NSF-sponsored experimental cluster MRI grant.
- Utilizing the campus Cyberinfrastructure Plan update process to identify and formalize the need for a specialized and dedicated cluster for sequence analysis. This process was a partnership with researchers, the campus Cyberinfrastructure Council, vendors, Research Computing Support Services, IRCF, Mizzou Advantage and the Office of Research to define, test, and fund the environment. The environment was purchased at the end of FY16 along with the other capacity investments at considerable savings, thanks to incentive funding by Dell.
- Building a teaching cluster from donated equipment to facilitate the training of students in scientific computing and utilization of High Performance Computing environments.
- Developing a secure (DCL4) research computing environment. This cluster will support researchers doing work with sensitive data such as patient records in a highly secure, standardized, and cost-effective manner.
- Increasing the number of researchers utilizing centralized research computing, storage and networking resources, the amount of computing core hours conducted, and the number of individuals trained in utilizing the environment.
- Negotiating a long-term sustainable 100Gbps connection to the next generation Internet2 Advanced Layer 2 System (AL2S) research network.

Enhanced Presence of Mizzou Research Computing

This past year, RCSS staff engaged with regional and national communities of support to continue to bring knowledge, experience, and expertise to MU. Activities included:

- Creating a state wide Cyberinfrastructure group (<http://ShowMeCI.org/>) to promote “Sharing Cyberinfrastructure information, education, and resources across the Show Me State.”
- Taking part in the NSF grant (Award #1620695) titled, *Advancing Research and Education through a National Network of Campus Research Computing Infrastructures – The CaRC Consortium*. This grant supports researchers utilizing advanced Cyberinfrastructure and helps fund a new “Cyberinfrastructure Engineer” position.

- Joining the national XSEDE Campus Champion Program <https://www.xsede.org/web/campus-champions> that provides national scale computing resources and support for scientific computing.
- Hosting MU's [Cyberinfrastructure Day](#) to promote the development and use of CI across campus where we presented a major campus [Cyberinfrastructure Plan](#).
- Exhibiting at [SuperCompute15](#), the preeminent High Performance Computing conference for scientific computing.

Improved Support for Researchers.

We are always striving to better support MU researchers with their computational needs. Current activities include:

- Keeping Research Computing Support Services web presence fresh with regular blog posts and service catalog updates.
- Offering weekly [Lewis training sessions](#).
- Expanding High Performance Computing training opportunities for researchers on campus.
- Hiring additional personnel to directly support researchers in their computational needs.

— Sep. 21, 2016

RCSS Spring 2017 Update

Lewis Cluster

The Lewis upgrade is now complete. The new hardware was in place last fall and by late January we completed the configuration and software installs (including more than 80 new scientific packages.) We are now focusing on users and how we can better serve their needs.

BioCompute Cluster

The BioCompute capacity of Lewis is now configured and online with a large number of bioinformatics software packages installed. Bi-weekly training sessions (in partnership with the IRCF) are now being held on Thursdays to help users get the most out of the cluster.

Teaching/Learning Cluster

The teaching cluster (tc.net.missouri.edu) is now available to students on campus for learning about research computing. This cluster has 11 compute nodes with between 96GB and 190GB of RAM. Instructors and students are welcome to use this as a part of their class with a number of restrictions. [Please read the policy](#).

In the next week or two, we will be having a short outage of Lewis to add five more investor nodes and to patch the cluster and make some optimizations to the SLURM resource scheduler. We will post an update to the Lewis announcement list when we have an exact date.

Secure4Cluster

We launched the Secure4 HPC environment. This cluster is for researchers who need to store/use Data Classification Level 3 and 4 (Health Insurance Portability and Accountability Act or HIPAA) data in a High Performance Computing environment. For more information, review the [service description](#).

High Throughput Computing Storage

We have added more than 600TB of new storage to the cluster for High Throughput Computing. This new storage service was launched for researchers requiring large, long-term storage for processing of large datasets where cost is a primary design criteria. This storage is only available inside the Lewis cluster or via rsync/ssh on the login nodes and is designed for large files/samples processed one-at-a-time. For example: DNA sequences, instrument images, video, etc. Storage pricing for FY17 is \$120/TB/60 months, paid in advance in increments of 10TB. For more information, review the [service description](#).

NSF MRI Grant

The [NSF MRI grant for HPC](#) out of the College of Engineering is nearing completion. This stand-alone cluster is being moved into production and we have been adding those nodes into the Lewis cluster so they will be accessible to all users via the General partition. This brings our total to more than 5,000

cores. We also migrated 10 GPU nodes (8 Nvidia K20m and 2 Nvidia K40 GPUS) into Lewis for general use.

[General Purpose Research Network \(GPRN\) /RNet 2.0](#)

The Research Network upgrade (RNet 2.0), Phase I, is nearing implementation. This will include the creation of the GPRN zone for researchers who have laptops and printers that need access to the Research Network. All Research Network ports will be moved to the campus networking ports (along with the associated port fee) in the next few months and will be completed before FY18. Phase II will be rolling out over the summer and will include enhanced performance and security as well as an update to our usage policies.

[Internet2/Great Plains Network](#)

The Research Network 100 Gbps Layer 2 connection has been moved to the GPN node in Kansas City for regional Layer 2 and Internet2 AL2S connectivity, leveraging our membership in the GPN to reduce operational costs. This connection will provide large dedicated bandwidth for connections to major data repositories such as [iPlant/CyVerse](#). Learn more about [the AL2S](#).

— Mar. 20, 2017

RCSS Supper 2017 Update	Sep. 27, 2017
--------------------------------	----------------------

Research Computing Support Services continues to expand grow. In August researchers consumed 2,932,217 core-hours of computation on the Lewis cluster up from 592,358 core-hours a year before.

In April we added five new nodes and integrated the remainder of the existing NSF MRI Research Cluster grant (NSF Award 1429294) into the Lewis cluster. This grant is ending after 3 years and the resources (around 70 nodes) are now a part of the Lewis cluster. This NSF grant was instrumental in the development of the Lewis environment as it is today and we are grateful to the NSF and the researchers in engineering that received the award.

The cluster now has over 5000 modern cores with an additional 400 older cores for interactive and debugging use. There are also now 10 GPU nodes (NVidia K20m or K40) from the MRI RC grant and we have five more GPU's (Nvidia P100 and four Nvidia GTX 1080 Ti's) that have been added to the cluster.

HPC Investment in the cluster continues to grow and we added 8 more investor nodes at the end of the summer. We now have a number of groups that have invested multiple times. They have been able to get started with a small investment and invest in additional capacity when they need it. Given the agility of our environment we have been able to make smaller purchases in-between larger investments for them. This has been made possible by a special relationship with Dell to get continuous pricing for our HPC investors allowing us to grow when and how we need to. This has the added benefit that we are now able to coordinate larger purchases by taking care of smaller immediate needs. We will be combining an increase to the community pool (available to all researchers at MU/UM) with a number of larger investments this fall for even deeper discounts.

Currently the cluster capacity is comprised of 20% individual investors, a 13% investment by the Office of Research and Mizzou Advantage specifically for Bioinformatics computations (BioCompute partition), 42% in an Engineering MRI grant [1] (NSF Award 1429294), and 25% by the Division of IT for the campus research computing community. In addition we have the equivalent of about 8% additional capacity (400 additional cores) of older cores for interactive jobs not included in previous count.

A recent example of a researcher immediately being able to utilize their investment demonstrates the power of investing. The researcher invested in 5 nodes and after migrating their workload onto the cluster they were able to run, in the first 41 days, 137,000 core hours on 512 of their 140 invested cores utilizing 125% of their investment during this period.

On July 1st we finished transitioning Research Network (RNet) connections in the building off old and slow equipment (some of which was over 10 years old) to the campus infrastructure. This allows us to provide RNet to all researchers across campus. In addition, we built a new General Purpose Research Network (GPRN) on campus that is a special zone that allows connectivity to and from research computing resources at high speed (Lewis for example) and at the same time provides greater protection from internet attacks. Research Network ports are available on any network port across campus and are charged the standard port fee. Experimental, dedicated, or high bandwidth capabilities on the Research Network are still available through partnerships with researchers, just reach out to our team if you are interested.

Improvement to the Research Network continues with the transition of our 100Gbps Internet2 AL2S connection that was part of a NSF CC*NIE grant [2] (NSF Award 1245795) to a dedicated connection to the Great Plains Network (GPN) in Kansas City. This connection provides 100Gbps Layer 2 connectivity to other GPN institutions and to other Internet2 AL2S sites. We are working on bringing IPv6 Internet2 connectivity (Layer 3) to this circuit as well. We are also upgrading our core cluster switch to 100Gbps which, in conjunction with the Layer 3 upgrades, will allow the cluster to connect to other universities at 100Gbps (Internet2 IPv6 sites) with a theoretical transfer rate of around 10GB/s.

The recently formed ShowMeCI.org group for “Sharing Cyberinfrastructure information, education, and resources across the Show Me State” has agreed to form a State Research Platform modeled after the Pacific and National Research Platform (PRP and NRP) (<http://prp.ucsd.edu/>). The goal of this effort is to enable researcher-to-researcher connectivity that spans the state and the region through a network of people enabled by standardized instrumentation and data transfer tools so that problems can be diagnosed and fixed. This effort is driven by researchers doing research across institutions with large data transfer needs, so if you have this need please contact us.

Finally, on the horizon is the replacement of the Lewis cluster storage with a high speed parallel filesystem (zLustre) supported by a 100Gbps core switch upgrade. This will be followed by another expansion of the Lewis cluster through HPC investment and nodes supported by RCSS.

For more information on the future of Cyberinfrastructure at the University of Missouri please attend CI-Day 2017 (<http://doit.missouri.edu/ci>).

References

1. <http://munews.missouri.edu/news-releases/2014/0926-nsf-grants-1-million-to-mu-to-expand-supercomputer-equipment-and-expertise-for-big-data-analytics-at-mu/>
2. <http://engineering.missouri.edu/2013/01/mu-researcher-secures-nsf-grant/>