The Institute for Cyber-Enabled Research: Regional Organization to Promote Computation in Science

Dirk Colbry, Bill Punch and Wolfgang Bauer
The Institute for Cyber-Enabled Research
Michigan State University
East Lansing, USA
{colbrydi; punch; bauerw}@msu.edu

***Abstract*— The Institute for Cyber-Enabled Research (iCER) at Michigan State University (MSU) was established in 2009 to coordinate and support multidisciplinary resources for computation and computational sciences. iCER is the home of MSU's centralized High Performance Computing resources, which include a heterogeneous compute cluster with various hardware designed to meet the needs of different scientists. The goal of iCER is not to maximize Flops, but instead to maximize the amount of quality science being accomplished. This paper outlines many of the initiatives that iCER has taken to support researchers at MSU and throughout Michigan; describes the structure and evolution of iCER; and offers insights for other institutions interested in putting together a similar computational organization.**

***Index Terms*—Cyber-Infrastructure, Human Support of eScience.**

I. Introduction: The Case for Computation

Established in 1855, Michigan State University (MSU) is one of the top research universities in the world. Home to nationally ranked and recognized academic, residential college, and service-learning programs, MSU is a diverse community of dedicated students and scholars, athletes and artists, scientists and leaders [1].

MSU has long been a leader in computational science, establishing a computer science program on campus in 1967 and a High Performance Computing Center (HPCC) in 2005. By 2009, researchers at MSU had completed approximately 450 cpu-years of computation, executing 1.8 million jobs on the HPCC. Ongoing internal evaluations of the HPCC pointed to a need for more than high performance computing (HPC) hardware in order to support advanced computation. We found that scientists also needed help establishing new scientific workflows and using advanced computing systems. The growing importance of computation for advances in natural science, engineering, medicine, and other areas led MSU to develop a new unit – iCER – to support computational research.

II. iCER Structure and Resources

In 2009, MSU established the Institute for Cyber-Enabled Research (iCER) as a subunit of the MSU Office of the Vice President of Research and Graduate Studies. iCER was created to support researchers and encourage collaborations across academia and industry, with the goal of improving the use of computation for research. iCER includes the previously existing HPCC, with a staff of three system administrators and a full-time administrative assistant working with direction from a computer science faculty member who serves as the HPCC Director.

iCER also supports a part-time HPC programmer (who also is a liaison to the Biology Department), and two full-time Computational Research Specialists to support students, faculty, and other computational researchers through education, training, and one-on-one consulting appointments. Currently, the iCER Computational Research Specialists include a PhD in computer science and a PhD in mathematics, both with experience applying computational techniques to research in a broad range of disciplines. In addition, iCER has established a large student internship program to train undergraduates, and supports a variety of post-doctoral scholars, computational specialists, and other experts to assist in research collaborations across campus. The current Director of iCER is a faculty member with a joint appointment in Physics.

III. Computation at Michigan State University

Currently at MSU there are 600+ registered iCER researchers, 50+ registered XSEDE users [2], and one Petascale Computing Resource Allocation (PRAC) award winner who will be running on Blue Waters [3]. Figure 1 shows the numbers of active users on the MSU HPCC since iCER was founded.

The hardware capacity of the HPCC now includes several thousand processor cores; hundreds of terabytes of high-performance, tiered data storage; large memory systems; a 10-gigabit per second Internet connection; and security, collaboration software, network management and virtualization platforms. More specifically, the HPCC currently includes:

General Purpose Beowulf Clusters

- A 384-core distributed memory cluster utilizing 2.2 GHz AMD Opteron 275 dual core processors connected by Single Data Rate (SDR) 10 gigabit Infiniband
- A 1024-core distributed memory cluster using Intel Xeon E5345 2.3 Ghz quad core processors, also connected by SDR Infiniband
- A 1536-core distributed memory cluster using Intel Xeon E5620 processors, connected by Quad Data Rate (QDR) 40 gigabit Infiniband

GPU Accelerated Systems

- Thirty-two 8-core graphics nodes utilizing 2.40GHz Intel Xeon E5530 processors. Each node has two NVidia Tesla M1060 General Purpose Graphics Processing Units
- An 8-core GPU accelerated server with 18GB of RAM. This node has two Nvidia Tesla C2075, one Nvidia Tesla C2070 and one Nvidia GTX 580
- A 4-core GPU accelerated server with 8GB of RAM. This node has one Nvidia GTX 480
- An 8-core GPU accelerated server with three Nvidia GTX 480s

Large Shared Memory Systems

- Four 32-core AMD Opteron-based multi-processor fat nodes each with 256 GB of available RAM and one 16 core AMD Opteron-based
- Two 32-core multi-processor fat node with Intel Xeon E7- 8837 processors and 512 GB of available RAM
- A 32-core multi-processor fat node with Intel Xeon E7- 8837 processors and 1 TB of available RAM
- Two 64-core multi-processor fat node with Intel Xeon E7- 8837 processors and 2 TB of available RAM

Condor Cluster

- ~200 node condor cluster with 2-4 cores and 2-6GB of memory per node

Storage

- High-speed parallel temporary file systems utilizing Lustre parallel file systems for fast file throughput
- Persistent mirrored storage of over two hundred terabytes

The MSU HPC system was purposely designed to accommodate a wide range of job types, in order to meet the varying needs of researchers. As demonstrated by Figure 2, during the first three years of iCER, system utilization has increased from 506 cpu-years (in 2009) to 1,741 cpu-years (in 2011). Figure 3 shows a similar increase in total number of jobs run, from 1.8 million jobs in 2009 to 2.4 million jobs in 2011.

Although both cpu-years and total jobs increased over the last three years, it is interesting to note that the total numbers of jobs dropped in 2010 (to 1.4 million jobs). This decrease is a reflection of users' increasing shift from running multiple, single-core jobs to running fewer, multi-core jobs. Consequently, the efficiency of our system utilization also increased during the first three years of iCER, as shown by Figure 4: in 2009, the average CPU utilization per job is about 1 core; in 2011, average CPU utilization increased to over 2.5 cores per job. It should be noted that cpu utilization does not include file I/O, and single-core jobs are still the vast majority (by four orders of magnitude).

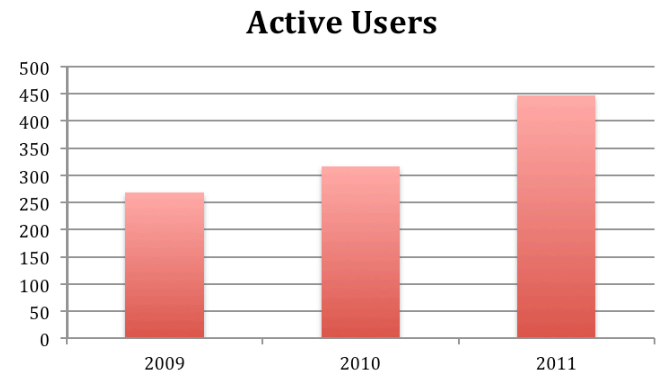

Figure 1 - Increase in the number of active users since 2009, when iCER was first formed.

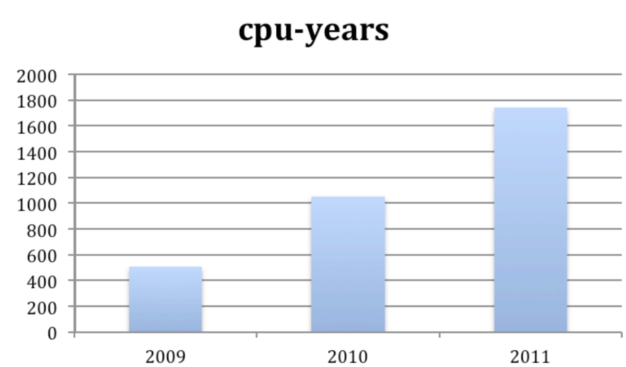

Figure 2 - Total cpu-years executed by HPCC users.

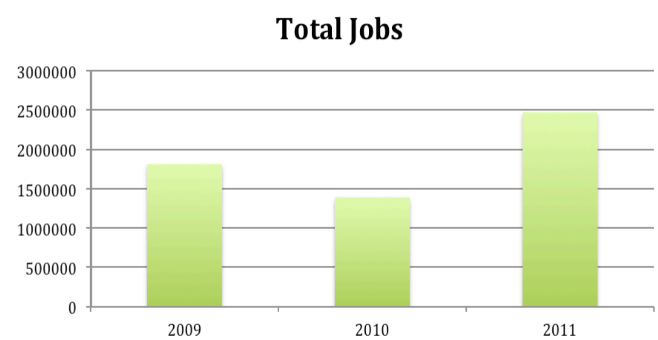

Figure 3 - Total jobs run by HPCC users.

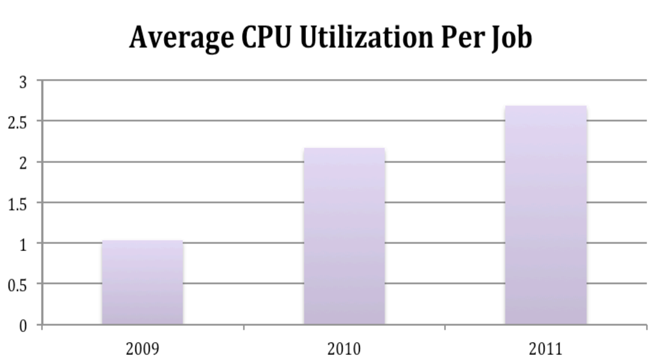

Figure 4 - Average CPU utilization per job on the HPCC.

IV. Decreasing "Mean Time to Science"

In addition to providing HPC hardware, the iCER team works hard to lower researchers' "mean time to science" by documenting and streamlining computational interfaces and reducing the barriers to entry for researchers unfamiliar with advanced computational hardware. For example, new users receive a "Portable HPC" USB flash drive that comes pre-loaded with ssh, X11 and SCP, and a plug-and-play format that allows Microsoft Windows users to gain immediate access to computational resources [4]. Developed by iCER staff, this "Portable HPC" drive allows novice users, with little or no prior HPC experience, to start using iCER resources quickly and easily. In addition to the general-purpose "Portable HPC," iCER has developed a variety of supports for domain-specific tasks. One example is providing pre-installed virtual machines for web-based biometric research, using software such as BLAST [5], Galaxy [6], and the UCSC Genome Browser [7].

Another way that iCER supports users and reduces the "mean time to science" is by developing a "one-queue" model for heterogeneous cluster scheduling [8]. In the one-queue approach, users specify what type of resources their jobs require, and the scheduler is tasked with determining which cluster resources best meet users' requests. The one-queue system is very simple for users, particularly for novices, since there is no need to parse queue names and descriptions to select the "best" resource for a specific job. In contrast, in traditional multi-queue systems, users must interpret the queue names ("short queue," "standby," etc.) and choose a queue to submit their job, then hope that they made the best choice to process the job quickly.

Although we feel that the HPC system available at MSU supports most of our researchers' needs, it is also iCER's goal to scale research to the point of exceeding the capabilities of local resources. For example, the HPCC staff has implemented an instance of a Globus Online Endpoint [9], which enables fast parallel file transfers to and from other national systems such as XSEDE [10] and Blue Waters [3]. iCER has also set up special accounts to make it easier for researchers to pay for EC2 [11] resources using internal university account numbers.

V. iCER Research Specialists

HPC resources are maintained by a team of system administrators and programmers, while iCER Computational Research Specialists focus on providing direct user support through task-specific training, continuing education and consulting services for researchers from all disciplines. iCER Research Specialists engage in a variety of tasks, as described in the remainder of this section.

A. Consulting

Computational Research Specialists provide consulting during office hours and scheduled appointments, working with undergraduates, graduate students, post-doctoral researchers, faculty and research staff members who need assistance with computational research projects.

iCER supports research in all disciplines, with recent projects including analyzing video data [12] and Parallel Time Integrators [13]. iCER Research Specialists work with the disciplinary researchers to evaluate the computational requirements of the research project and develop solutions for research problems. Depending on the nature of the research and the computational requirements, the iCER Research Specialist will provide or develop a solution for known computational issues, or may develop a research project focusing on a novel computational challenge related to the disciplinary research.

B. Testing, Troubleshooting and Benchmarking

iCER Research Specialists frequently help disciplinary researchers evaluate available software and hardware and select appropriate solutions. iCER Research Specialists install software on HPC systems, write tests scripts, troubleshoot system problems and run benchmarks to evaluate the performance of algorithms on different hardware configurations. Research Specialists also assist HPC system administrators in monitoring the performance and stability of the HPCC systems.

C. Debugging and Programming

iCER's Computational Research Specialists frequently assist disciplinary researchers with debugging problems that arise when trying to compile research-related programs on HPCC resources or link to HPC-specific libraries. C, C++, Python, FORTRAN, MATLAB, bash, Perl, R, openmp, mpi, pthreads, and cuda are used most frequently, but iCER Research Specialists work with other languages and programming environments as well. On occasion, Research Specialists are asked to help disciplinary researchers write code that will optimize or parallelize an existing application.

D. Training, Mentoring and Education

iCER Research Specialists teach in a variety of formats, including developing instructional and reference materials for in-person and online delivery. Examples of recent educational topics include algorithm parallelization techniques, training on how to use specific software and hardware resources, and methods for software carpentry. Training formats include one-on-one tutorials, seminars and for-credit classes.

E. Grant Writing

iCER Research Specialists help educate disciplinary researchers about computational resources and provide feedback on appropriate resources to include in various funding proposals. iCER Research Specialists also participate in grant development as investigators, co-investigators, senior personnel or consultants, as appropriate to the project.

F. Outreach and Assessment

Research Specialists participate in campus and community outreach activities, including developing and providing educational programs related to computational science and iCER resources. In addition, Research Specialists develop communications materials (research highlights, iCER website, news articles, etc.) and tools for evaluating and improving iCER research and outreach activities.

VI. User Training, Education and Support

Part of iCER's mission is to provide opportunities for user training and research education, and to support ongoing research collaborations. Computational Research Specialists at iCER provide extended, one-on-one support to disciplinary researchers. Frequently, this assistance includes identifying appropriate computational tools and resources; improving efficiency (parallelizing experiments, porting code to HPC systems, etc.); and identifying and training students and programmers who can provide long-term assistance for computational projects.

iCER provides a number of educational seminars, classes and programs for researchers, faculty, and students. Topics from recent iCER presentations and seminars include:

- Software Carpentry [14]
- MATLAB
- Bioinformatics
- MPI Programming
- OpenMP
- GPGPU Programming

In addition to providing regular computational research training for members of the MSU community, iCER promotes collaborative research by hosting regional/national research seminars and workshops focused on computation. For example, in 2011 iCER hosted an ENZO developer workshop where 22 participants from UC San Diego, Michigan State University, and the National Center for Supercomputing Applications at the University of Illinois in Urbana-Champaign gathered to report current progress and discuss future code development. There were also discussions about grant funding among the PI-level attendees, which resulted in submitted NSF RFPs. Two of the visitors also gave seminars about ENZO to the broader MSU community.

For the last three years, iCER has been a host site for the annual Summer Virtual School in Computing and Engineering (VSCSE) [11]. Hosted classes include; Big Data For Science (2010), Petascale Programming Environments and Tools (2011), Proven Algorithmic Techniques for Many-core Processors (2011, 2012) and Science Cloud Summer School (2012). VSCSE participants have come from MSU, the University of Michigan, and Lansing Community College.

iCER provides travel funding to assist collaborators from other locations in traveling to campus to work with MSU's computational researchers. iCER also sponsors an annual seminar series that brings computational researchers from across the country to MSU. In addition, iCER has jointly supported a number of post-doctoral scholars who are working on computational research projects in various disciplines on campus.

VII. Undergraduate Research and Internship Program

iCER is committed to training the next generation of computational scientists and system administrators, and has developed an internship program for undergraduates that employs an average of 10 students during the academic year, and 15 students during the summer months. These students work in a variety of areas, from hardware maintenance to computational research, and the iCER undergraduate intern program provides a unique experience for students to learn practical skills and gain responsibility and experience over time. iCER student interns assist MSU research scientists with computational projects, gain valuable system administration experience, and learn to install, evaluate and teach others about various high performance software. A sample of recent student projects include:

- Evaluating and deploying CONDOR for high throughput computing
- Visualization of MSU's machine room status [15]
- WIKI installation and migration
- Developing and improving web interfaces to support user interactions with iCER staff
- Creating a YUM repo server
- Node maintenance (hardware trouble shooting, replacement, burn in tests, etc.)
- Installing, testing and documenting Makeflow, a workflow engine developed at Notre Dame for executing large workflows on clusters
- Installing, testing, and documenting parallel profilers such as Tau and Scalasca [16]
- Installing, testing and documenting visualization software for Supernova Simulations using paraview and OSMesa [17]
- Assisting in parallelizing serial simulation code using OpenMP and HDF5
- Installing, testing and documenting the use of the eclipse parallel toolkit
- Integration and development of a series of tools (called Powertools) developed by iCER staff and users [18]

iCER staff also developed an undergraduate honors research seminar for first- and second-year students at MSU. The goal of this 3-credit course is to introduce novice students (with little or no prior research experience) to both the process and content of research in computational science. Students meet weekly throughout the academic year to learn about various topics in research and computation, and develop individual research projects to present at the University's undergraduate research forum. In 2010, six undergraduate from various disciplines (Astrophysics, Mathematics, Education, Electrical Engineering, Packaging, Supply Chain Management) participated in the honors research seminar, which examined ways to improve the efficiency and environmental sustainability of the HPC server room [15]. This honors research seminar is being offered again in 2012, with a focus on image processing and computational research.

VIII. Community Outreach

Part of iCER's role is to expand public understanding of the importance of computation in the advancement of science. In addition to seminars and training sessions for the MSU community, Computational Research Specialists from iCER regularly conduct outreach activities within the broader community. For example, in spring 2012 iCER employees visited several computer science classes at Anchor Bay High School in Ira, Michigan (a northern suburb of Detroit). The students were given a brief presentation about high performance computing and the role of computation in scientific research, and then they had the opportunity to take apart several compute nodes from a retired HPC cluster. This hands-on activity allowed students see the "insides" of a computer and learn how they functioned – one student group was particularly thrilled when they were able to successfully re-assemble and power up the node they had taken apart.

Researchers associated with iCER have also developed a number of courses for the MSU Evening College, which offers non-credit personal development classes to community members [19]. Recent Evening College offerings include a course outlining advances in computational science (and a tour of the HPC server room), and a “Coffee with the Profs” event where community members have the opportunity to meet senior faculty in a conversational setting.

IX. Regional Partnerships

In 2010, MSU partnered with Central Michigan University (CMU) to expand the HPC capacity of iCER and make high performance computing more accessible to researchers at both Universities. Researchers at MSU and CMU use iCER resources for diverse projects, including climate modeling [20]; Galaxy Formation [21]; genetic programming [22]; and numerical simulations that help explain experimental observations at MSU’s Facility for Rare Isotope Beams [23].

As a founding member of the Great Lakes Consortium for Petascale Computing (GLCPC) [23], MSU hosts some of the Consortium’s Summer School classes to support researchers interested in working with petascale computation. iCER research faculty have also participated extensively in the Blue Waters Undergraduate Petascale Education Program, helping mentor 3 undergraduate students in the last two years.

X. Research Impacts

Since iCER was founded in 2009, more than 5.5 million jobs have run on MSU HPC resources, utilizing more than 3,300 cpu-years. In addition, iCER’s Computational Research Specialists have completed more than 250 individual consulting appointments; resolved more than 6,000 user support tickets; and assisted in training and placing 30 students and programmers with research projects across campus. With support from iCER staff and computational resources, at least 175 documented scientific papers have been published. (An updated list of many iCER-related papers, projects and grants is available online at: http://icer.msu.edu/ [24])

External grant support is the gold standard by which all universities measure their research quality, and one explicit goal in establishing iCER at MSU was to increase our computational researchers' competitiveness for external grants. To this end, iCER provides 50% matching towards a postdoc for large external multidisciplinary or multi-investigator grants. While it is, of course, hard to quantify the impact of any single component in the overall success of a grant proposal, it is still clear that external grant support for computational sciences has sharply increased at MSU since the founding of iCER.

Several large group grants that included iCER 50/50 postdoctoral position matches have been funded. Among these is the NSF Cyber-Enabled Discovery and Innovation (CDI) grant "Modeling and Data Analysis Initiative" collaboration (MADAI) [25]. Perhaps our most important success is the NSF Science and Technology Center (STC) BEACON, which studies evolution in action through computer simulation [24]. BEACON and iCER share office space in a dedicated facility, which enables close integration and a high degree of leveraging of MSU resources. Senior personnel in iCER, including the present authors, are charged with initiation and/or supporting these and other grant opportunities as part of iCER's mission.

Another extremely important contribution of iCER to the success of the externally funded research programs at MSU is the "HPCC buy-in" program. This program enables grant-supported researchers to utilize 100% of their available hardware funding for purchasing computing hardware, without the need to budget for associated overhead expenses (hardware and software maintenance, power, cooling, and so on). Once a researcher commits funds for hardware purchases, these nodes are integrated in the general pool of resources of the HPCC. These dedicated resources are available to the purchaser at any time through priority queue access, but any surplus capacity is made available to all users. This tiered access model is made possible, in part, through a four hour guaranteed access window and our use of the Berkley Lab Checkpoint Restart (BLCR) which enables lower priority jobs to move to alternative resources when buy-in users need access to their purchased nodes [26]. Obviously, this model greatly leverages external research funds, but it also benefits the wider community of iCER's HPC users.

XI. Discussion and Lessons Learned

The Institute for Cyber-Enabled Research at MSU was established in order to provide broad support for computational research. Rather than just providing advanced HPC hardware, the goal of iCER is to maximize the quantity and quality of computational research on campus, and assist the MSU community in leveraging regional and national resources to support large-scale research and discovery.

iCER has enjoyed many successes in its first three years, including a dramatic increase in total users and cpu-years; extensive education and outreach programs; and research collaborations on campus, throughout the Midwest, and across the nation. Key elements that contributed to the success of iCER include:

- Investing in HPC hardware, computational software and system administration personnel
- Investing in Computational Research Specialists to provide user support and education
- Developing extensive education and training programs for computational researchers in various disciplines
- Reducing barriers to entry for novice users of HPC systems
- Training undergraduates and graduate students in advanced computation, and matching skilled student programmers with disciplinary research projects
- Providing cost-sharing to support programmers and post-doctoral researchers who serve as liaisons between iCER and disciplinary research projects
- Partnering locally, regionally and nationally to provide computational resources and training
- Developing outreach activities about computational research to educate and engage the public, including high school students and early undergraduates

iCER has also encountered some challenges in its first three years. For instance, computational research comes in many varieties and it can be difficult to develop training and support programs that meet the wide range of user needs. It can also be difficult to identify and hire individuals with appropriate skills for a position like "Computational Research Specialist." The job description for this position at iCER calls for someone with an earned Ph.D. in a computational area and experience with interdisciplinary research. iCER's Computational Research Specialists participate in teaching, research, and grant-writing, but are neither tenure-track faculty nor traditional post-doctoral scholars.

As the practice of computational research evolves, there is an emerging need for experts in computational research who have the ability to communicate and collaborate in multidisciplinary settings. Training programs like the Campus Champions [24] and new positions like the Computational Research Specialists at iCER are an early effort to meet the growing need for computational research support, but the field is still emerging and the career pathways are not yet clear.

Acknowledgment

The authors would like to acknowledge the iCER team members: Dr. Ben Ong, iCER Research Specialist; Andrew Keen, system administrator and lead architect of the HPCC; Jim Likert, system administrator, designer of the iCER usage database; John Johnston, biology liaison, and lead of the iCER Bioinformatics effort; Greg Mason, system administrator and architect of HPCC file systems including Globus Online; and Kelly Osborn, administrative assistant and developer of the EC2 bridging program.

References

[1] "Michigan State University (MSU). Est. 1855. East Lansing, Michigan USA." [Online]. Available: http://www.msu.edu/.

[2] "XSEDE, Extreme Science and Engineering Discovery Environment, supported by National Science Foundation grant number OCI-1053575." [Online]. Available: https://www.xsede.org/.

[3] "About the Blue Waters project." [Online]. Available: http://www.ncsa.illinois.edu/BlueWaters/.

[4] D. Colbry, "Reducing the barrier to entry using portable apps," in *Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery*, New York, NY, USA, 2011, pp. 47:1–47:1.

[5] "BLAST: Basic Local Alignment Search Tool." [Online]. Available: http://blast.ncbi.nlm.nih.gov/Blast.cgi.

[6] "The Galaxy Project: Online bioinformatics analysis for everyone." [Online]. Available: http://g2.bx.psu.edu/.

[7] "UCSC Genome Browser Home." [Online]. Available: http://genome.ucsc.edu/.

[8] A. Keen, "Lessons Learned When Building a Greenfield High Performance Computing Ecosystem," presented at the 26th Large Installation System Administration conference (LISA), San Diego, California, USA, 2012.

[9] "Globus Online | Reliable File Transfer. No IT Required." [Online]. Available: https://www.globusonline.org/.

[10] "GFFS – Global Federated File System browse." [Online]. Available: http://genesis2.virginia.edu/wiki/Main/GFFS.

[11] "Amazon Elastic Compute Cloud (Amazon EC2)." [Online]. Available: http://aws.amazon.com/ec2/.

[12] N. Ingle, T. Door, D. Colbry, and F. Dyer, "Coordination of Vision and Action in Chameleons," presented at the 49th Animal Behavior Society Annual Meeting,, Albuquerque, NM, 2012.

[13] A. Christlieb, "Implicit Parallel Time Integrators," *Journal of Scientific Computing*, vol. 46, no. 1, 2010.

[14] G. Wilson, "Software Carpentry." [Online]. Available: http://software-carpentry.org/.

[15] A. Addis and A. Keen, "Visualization of Computing Cluster Data," presented at the XSEDE Conference, Chicago, IL, 2012.

[16] "VSCSE - Virtual School of Computational Science and Engineering." [Online]. Available: http://www.vscse.org/.

[17] X. Peng and D. Colbry, "Visualizing Simulated Supernova," presented at the MSU Summer Undergraduate Research Forum (SURF), East Lansing, MI, 2011.

[18] R. Picket, Colbry, Dirk, and I. Sagert, "Computational science education through multiple approach parallelization for supernova simulation development," *Submitted to Journal of Computational Science Education*, 2012.

[19] "Alumni Lifelong Education Evening College | MSU Alumni Association." [Online]. Available: http://alumni.msu.edu/eveningcollege/.

[20] N. Moore, "East African Food Security as Influenced by Future Climate Change and Land Use Change at Local to Regional Scales," *Climatic Change*, vol. 107, 2011.

[21] Gómez, "Signatures of minor mergers in Milky Way-like disc kinematics: Ringing revisited," *ArXiv e-prints*, 2011.

[22] Jensen, "On the Use of Genetic Programming for Automated Refactoring and the Introduction of Design Patterns," 2010. .

[23] D. Colbry and K. Luchini-Colbry, "CyberGreen: Hands-On Engineering Research in Sustainability and Supercomputing"," presented at the ASEE North Central Section 2012 Conference, Ada, Ohio, 2012.

[24] Michigan State University, "Institute for Cyber Enabled Research (ICER)." [Online]. Available: http://icer.msu.edu/.

[25] NSF, Duke, and U North Carolina, "Models and Data Analysis Initiative (MADAI)." [Online]. Available: http://madai.msu.edu/. [Accessed: 22-Mar-2010].

[26] P. H. Hargrove and J. C. Duell, "Berkeley lab checkpoint/restart (blcr) for linux clusters," in *Journal of Physics: Conference Series*, 2006, vol. 46, p. 494.