UNIVERSITY OF MICHIGAN
RESEARCH CYBERINFRASTRUCTURE (CI) COMMITTEE

REPORT TO THE VICE PRESIDENT FOR RESEARCH

AND

THE VICE PROVOST FOR ACADEMIC INFORMATION

December 24, 2007

Committee Members:
Brian Athey, Chair (UMMS, CCMB)
Sharon Glotzer, Co-Chair (CoE)
Ulysses Balis (UMMS)
Tom Finholt (SI)
Myron Gutmann (LS&A, ISR)
Ted Hanss (UMMS)
Farnam Jahanian (CoE)
Paul Killey (CoE)
Alexey Kondrashov (LS&A, LSI)
Shawn McKee (LS&A)
Homer Neal (LS&A)
Gil Omenn (UMMS, CCMB)
Richard Rood (CoE)
Ann Marie Sastry (CoE)

UNIVERSITY OF MICHIGAN
RESEARCH CYBERINFRASTRUCTURE COMMITTEE (R-CIC)

REPORT TO THE VICE PRESIDENT FOR RESEARCH
AND
THE VICE PROVOST FOR ACADEMIC INFORMATION

Table of Contents

1.0 Executive Summary

This report represents a consensus view of the University of Michigan Research Cyberinfrastructure Committee (R-CIC). The R-CIC is made up of senior faculty and IT leaders who were chosen based on their acknowledged national expertise and prominence in CI, university administration, and IT leadership.

1.1 Key Findings of the R-CIC:

U-M History and Roles: The University of Michigan has an illustrious history of computing and information science leadership going back to the mid-1950s. U-M is the birth-place of the higher education Internet and early distributed research computing, and has been a major cyberinfrastructure creator and thought leader.

Central Role of CI to Enable Future Research at U-M: Cyberinfrastructure is critical for all present and future fields of research that represent the activities of the 19 schools and colleges at Michigan and its many fine institutes and centers. This spans the natural sciences, engineering, social and political sciences, arts and humanities, and law, policy and business research domains. The R-CIC finds that the University of Michigan administration does not consider research CI a "mission-critical" institutional priority.

U-M Research CI Strengths: U-M has great elements of strength and innovation in research CI as well as CI research and implementation, but they are dispersed, disconnected and under-leveraged. As a result, U-M has a significant set of isolated flagship efforts that are at significant risk of being unsustainable. Coordinating and leveraging these CI strengths is critical to future U-M research competitiveness, U-M leadership in CI, and emerging CI-enabled fields.

The Data Explosion: Challenges relating to the creation, preservation, use, and sharing of digital data are exploding at an alarming rate. A strategic plan is needed that leverages the broad set of strengths in the campus academic units, university libraries, the U-M Institute for Social Research (ISR) and its Inter-University Consortium for Political and Social Research (IPCSR), IT units, DRDA, the Health System, and the Office of the General Counsel. Regulatory compliance with data-sharing and reporting requirements from sponsoring agencies is an issue demanding immediate attention. Likewise, back-up and security must be addressed campus-wide.

Research CI Facilities and Resources Required: A well-managed and adequately resourced set of integrated campus-wide CI service facilities is critical to faculty research productivity, and to faculty recruitment and retention efforts. Agreed-upon policies for provision and support of CI resources for faculty are needed. The elements of shared research CI combine facilities, computing systems, databases and archives, storage systems, software and algorithmic codes, visualization systems, advanced instrumentation, collaborative tools, and trained personnel – all linked by high-capacity Internet-based communication networks across campus and to appropriate outside research networks and commercial Internets.

The Michigan Academic Computing Center (MACC): The MACC is suffering from a lack of leadership and direction that may become a crisis without immediate intervention. There is no clear plan for unit participation beyond a simple "condo" model, which allows for minimal sharing or leveraging of personnel or expertise. There is no process by which a shared participatory model can be established for defining and delivering CI services via the MACC (storage, cycles, etc.). The current management of the MACC by non-research IT units is insufficient for successful implementation of the MACC as a campus-wide research resource. If left on its current course, the MACC will fail to fulfill its original mission. Moreover, the MACC is projected to be at capacity within two years, requiring the identification of additional machine room space. This issue requires urgent action.

Research CI at Comparable Institutions: The committee examined CI activities at many comparable and smaller research institutions such as Indiana University, Pennsylvania State University (PSU), Purdue University, University of Illinois, the University of Utah, Clemson University, Cornell University, Harvard University, Rensselear Polytechnic Institute (RPI), University of Pittsburgh, and the University of Southern California (USC). Each of these institutions has established a more extensive set of CI resources for supporting research, with a deeper level of institutional commitment, than the U-M.

Deferral of Investment in Research CI: There has been a long-term deferral of “research IT” strategy and coordinated investment at U-M, from central and unit administration perspectives, going back at least 10 years. This deferral has left schools, colleges, institutes, centers, and individual researchers to “fend for themselves”, in the absence of any strategic plan for research CI. The result of this extreme decentralization is that U-M has a severely under-resourced set of CI capabilities for researchers, far below similar and even smaller research institutions.

Scope of “Mission-critical” U-M IT Organizations: In contrast to research CI, several administrative IT units, with appropriately large operating budgets, are considered “mission-critical” to the U-M. These include Information Technology Central Services (ITCS, $25M/yr); Michigan Administrative Information Systems (MAIS, ~$25M/yr); and the Medical Center Information Technology unit (MCIT, ~$50M/yr). Those three units have an aggregate base budget of $100M, and an aggregate of 1250 base FTEs, not including major initiatives. (Numbers provided by the Office of the Vice Provost for Academic Information.)

Risks of Inaction: To the extent that the U-M is not investing in research CI, it will be forfeiting the opportunity to be future leaders in research — all to the detriment of the biomedical and life sciences, engineering and the physical sciences, the natural sciences, the social and political sciences, education research, and 21st century endeavors in the humanities and the arts.

2.2 <u>*Key Recommendations of the R-CIC:*</u>

Importance of Research CI to U-M: The University of Michigan executive officers must recognize the rapidly growing importance of CI to the university research mission, and act boldly to maintain our status as a leading research university in the 21st century by investing in research CI. The CI initiatives upon which U-M must capitalize are listed below; these efforts will be costly and take hard work. They will, however, pay enormous benefits to the university by further lifting the capacities and national prominence of our entire research base - both individual investigator-initiated research, as well as team-based interdisciplinary research efforts. They will also greatly enhance and enable the instructional enterprise.

Establishment of an Office of Research Cyberinfrastructure: U-M should establish an Office of Research Cyberinfrastructure (ORCI) within the Office of the Vice President for Research (OVPR). The ORCI would have the mandate and budget to lead the creation of a federated CI system for research that spans the University of Michigan. The ORCI would provide OVPR with clear institutional planning, budget recommendations, and proposed integration of existing and future large-scale, interdisciplinary campus CI activities. The ORCI will create institutional priorities, and be responsive to the opportunities provided by state and federal funding agencies and programs. The ORCI should be the primary internal and external point-of-contact and coordination for all research-CI inquiries and new large-scale university activities.

Creation of a CI Operations Center: U-M should establish a well-supported and highly visible Cyberinfrastructure Operations Center (CIOC) within the Office of the Vice Provost for Academic Information to coordinate efforts among ORCI and existing IT organizations on campus, including ITCS, CAEN, and MSIS, to provision IT services where appropriate. CIOC should also coordinate with the Libraries and ORCI in support of research CI.

Computational Science Faculty Recruitment and Retainment: U-M should establish at least fifteen partially-funded or endowed instructional- and research-track lines for faculty in emerging CI areas, including petascale computing and data science. Coordination across departments and/or units should be encouraged to recruit "cyber" faculty whose cutting edge, high performance computing-based and data science-based research spans computer science and/or information science, and one or more application domains such as energy or nanoscience. Such hires are needed to maintain cutting-edge research at U-M, and ultimately to move U-M ahead of our peer institutions in the area of cyber- enabled discovery and innovation and position us for leadership in this emerging, cross-cutting area. This is a critical need that must be addressed.

Expansion and Repurposing of the Center for Advanced Computing: U-M should expand and re-purpose the Center for Advanced Computing (CAC), currently run by the College of Engineering. The new center, CAC++, should be developed as a campus-wide center for high performance computing that is a user-focused service provider, serving all participating schools, colleges and units, and overseeing the MACC facility and other federated research CI resources and machine rooms. The CAC++ functions should include acquisition support, housing and system administration of all high performance research computers and large research data servers, including faculty-acquired research clusters and other platforms; provision of open-access compute cycles for research to campus-wide users; management and oversight of the MACC; development and implementation of a sustainable, cost-neutral plan for long-term hardware/software acquisition, development and support; assistance, through the use of expert HPC consultants, with code development and implementation for faculty-led research projects; and applied research to test research software and hardware to optimize its effectiveness in use.

Petaflop Scale Computational Science: The U-M should seek to be a major player in petascale science, which is at the cutting edge of High-End Scientific Computing, Modeling, Simulation, and Data Analysis. Good progress toward this goal has been made recently, relating to key U-M participation in the recent $208M petascale system award to the National Center for Supercomputing Applications (NCSA), based at the University of Illinois Urbana-Champaign (UIUC). U-M was a founder with NCSA of the Great Lakes Consortium for Petascale Computation (GLCPC). As such, U-M now needs to maintain its enviable leadership positions in the GLCPC and its Virtual School of Computational Science and Engineering, both based at NCSA. U-M must, therefore, seriously invest in research CI to take advantage of this position. As such, the U-M will also need to establish significant, shared, cutting-edge, 10-100 Tflop computing clusters (Tier 3 in the NSF hierarchy) based at the Michigan Academic Computing Center (MACC) to credibly participate. It will also require a set of linked, first-class, visualization facilities to properly interface our researchers with the petascale systems.

Petabyte Scale Data and Analysis: U-M has been a leader in large-scale data handling, analysis, and networked transmission in its work in the high-energy physics community (US-ATLAS). This pioneering effort needs to be expanded broadly to all fields, building expertise and capacity in data organization, data mining and visualization in all branches of science, especially in the life sciences and biomedical fields. As one example, the implications of terabytes per week of human genome sequence data for research and eventual clinical use is a sobering example of what we are facing as an institution and as a nation as personalized medicine comes of age. This and other examples means we need to develop high performance petabyte storage and archival systems and place them on national e-Science grid portals. In addition, the U-M needs to seek several focused public/private partnerships to house and transmit datasets of national and international interest. This work should be tightly coordinated with ongoing and future efforts with the University Library, in particular within the context of our Google partnership.

Coordinated Access to National CI Resources: U-M should assist researchers in accessing a coordinated set of national computational resources, providing access to Tier 1 and 2 resources. This coordination function would work with campus units and researchers to access computing cycles from national partners such as the NSF petascale initiative at NCSA, the NSF Distributed Terascale Facility (DTF), which includes the San Diego Supercomputer Center (SDSC), Pittsburgh Supercomputing Center (PSC), and the Texas Advanced Computing Center (TACC); NIH Resources such as the National Biomedical Computing Centers, the Biomedical Informatics Research Network (BIRN), the CTSA Informatics Consortium, the cancer BioGrid (caBIG); DoE resources, including national laboratory sites and NERSC; and DoD computing facilities at the services laboratories and other sites. In addition, the U-M should be aggressive in establishing partnerships with IBM, Sun, and other computer vendors to facilitate researcher access to compute cycles. Such access is not a substitute for substantial U-M investment in local computational resources, but is necessary to provide a well-balanced portfolio of modern resources to the U-M research community.

Virtual Organizations (VO) and Distributed Science Communities: This domain is a current strength of U-M and should be an integral part of all interdisciplinary projects. The School of Information (SI) has been a noted thought leader in this area, and should be incentivised to further build in this area. In addition, U-M needs to seek to be the supplier of key VO capabilities for individual investigator and team-based virtual research initiatives and projects, by creatively building on its long-term investments in Sakai, C-Tools, and M-SCRIBE systems. An example of a prominent U-M research group building modern portal technology, leveraging the U-M Sakai investment, is the NIH National Center for Integrative Biomedical Informatics (NCIBI) in the U-M Center for Computational Medicine and Biology (CCMB). A much more vital research portal activity is desirable, and could be accomplished by building upon expertise at SI and other units.

Large-scale automated multimedia recording and archiving of classroom lectures, as was demonstrated in the M-SCRIBE pilot research project, is not just a student study aid. It is a cutting-edge activity that will fuel many new pedagogical research studies. Similar recording systems are already being implemented on a large scale at several major universities such as Stanford and Duke. In addition, collaborative tools are an integral part of the conduct of team-based research projects, and contributions to the advancement of this increasingly important area through efforts such as the Michigan ATLAS Collaboratory Project should be encouraged and supported.

Education and Workforce Development: The U-M research and instructional missions are inextricably linked, and advances in cyberinfrastructure are equally necessary in the instructional arena. Increasingly, modeling, simulation, prediction, and data intensive mining of knowledge are critical to many types of knowledge-based activities — both discovery and learning. We need to train our students in contemporary and emerging methodologies and ways of working. Thus, the investments and activities proposed by the R-CIC should have additional benefit to the instructional mission of the U-M. As such, it is critical that the U-M recognize and capitalize on this relationship, and closely coordinate research CI efforts with the instructional CI efforts being led by the office of the Vice Provost for Academic Information. The committee also recommends a more serious effort in outreach and dissemination using CI, including upgrading our capabilities and presence in the *Research Channel* television network.

Exploring Creative Industry Partnerships: New commodity computing cycle and data server paradigms offered by Amazon and Google are likely to be game-changers for research computing. The U-M should be aggressive in establishing partnerships with these companies (in particular, leveraging U-M's existing Google partnership) to facilitate researcher access to compute cycles and data. Success in this endeavor will have wide-ranging and substantial impact on computing-enabled research at U-M, and could mean substantial savings for Michigan in terms of computer housing and maintenance.

Level of Investment Required: The committee tried but was unable to acquire research CI expenditure information on individual schools and colleges. Thus, an aggregate understanding of current research (and instructional) CI expenditures at U-M is unavailable, and must be urgently compiled by the Office of the Vice Provost for Academic Information. The committee did find, however, that several schools and colleges are investing significantly in research CI; and identified an aggregate of several millions of dollars of annual research CI investments made by the Provost's Office, UMMS, CoE, and LS&A, and other schools and colleges.

The magnitude of investment required, from central administration and unit sources, should be commensurate with mission-critical IT organizations, and should strategically leverage these IT units wherever possible. A significant additional aggregate investment, well in excess of ten million dollars annually, will be required. The mission-critical IT units such as ITCS, MAIS, and MCIT provide excellent examples of scale.

2.0 Committee Charge and Process

The University of Michigan Research Cyberinfrastructure Committee (R-CIC) was convened in November 2006, by the Vice President for Research, Stephen Forrest, and the Vice Provost for Academic Information, John King. R-CIC co-chairs, senior faculty, and IT administration leaders were chosen based on their acknowledged national expertise and prominence in CI and research administration and IT leadership. The overarching charge to the committee was to provide an assessment of the importance of cyberinfrastructure to the future of the U-M research mission and its competitiveness. The committee was also charged with assessing the current U-M CI strategic plan broadly, focusing primarily on current strengths and weaknesses in CI leadership and activities in support of the research mission at U-M, drawing comparisons with other similar research institutions where appropriate. In addition, the committee was urged to provide guidance and insight on how to best position U-M to be a future leader in regional, national, and international research CI.

The committee met formally three times and also had numerous off-line discussions with key U-M faculty engaged in research and campus IT leaders. R-CIC co-chairs Brian Athey and Sharon Glotzer had many discussions with VP Forrest, Vice Provost King, and key thought leaders at U-M and nationally throughout the process. Athey and Glotzer also met with computing and IT committees at the UMMS and CoE, and had discussions with the CoE Center for Advanced Computing (CAC). They also met with several members of the MACC Operating Committee. Athey had contact with several of the Research IT groups at the Dental School, College of Pharmacy, and the School of Public Health.

3.0 Detailed Findings

Listed below are the detailed findings of the R-CIC.

3.1 The main finding of the R-CIC is that the University of Michigan administration does not consider research CI a "mission-critical" institutional priority, to the detriment of our current and future research standing and competitiveness in CI-enabled and CI-driven fields.

3.2 The U-M has fallen alarmingly behind peer institutions in addressing the challenges and opportunities afforded by CI as a key enabler of 21st century team-based interdisciplinary research (IDR), putting U-M at risk for maintaining our present research standing and capturing major new sponsored CI-based initiatives.

3.3 The U-M has great elements of strength and innovation in research CI and CI research and implementation, but they are dispersed, disconnected, and under-leveraged. As a result, U-M has a significant set of isolated flagship efforts that are at risk of being unsustainable. Coordinating and leveraging these strengths is critical to U-M research competitiveness and U-M leadership in CI and emerging CI-enabled fields.

3.4 There has been a long-term deferral of "research IT" strategy and coordinated investment at U-M from central and unit administration perspectives, going back at least 10 years. This has left schools, colleges, institutes, centers, and individual researchers to "fend for themselves" in the absence of any strategic plan for research CI. The result of this "extreme decentralization" is that U-M has a severely under-resourced set of CI capabilities for researchers, far below similar and even smaller research institutions.

3.5 A well-managed and adequately resourced set of integrated campus-wide CI service facilities is critical to faculty research productivity, and to faculty recruitment and retention efforts. Agreed-upon policies for provision and support of CI resources for faculty are needed.

3.6 There has been a severe lack of strategic planning for interconnecting the existing machine rooms on campus to support research and promote resource sharing and security. The availability of coordinated and integrated machine rooms to support cluster-based computing and server-based data needs for faculty and student research endeavors is an essential need for modern research. Broad input from, and oversight by, the U-M research community is essential but has been lacking.

3.7 The newly built Michigan Academic Computing Center (MACC) is suffering from a lack of leadership and direction that may become a crisis without immediate intervention. There is no clear plan for unit participation beyond the "condo" model. There is no process by which a shared participatory model can be established for defining and delivering CI services via the MACC (storage, cycles, etc.). The current management of the MACC by non-research IT units is insufficient for successful implementation of the MACC as a campus-wide research resource.

If left on its current course, the MACC will fail to fulfill its original mission. The MACC was built to address the acute need for conditioned machine room space to accommodate current requirements for, and expected rapid growth of, research computing, data server, and networking capacity. As currently implemented, the MACC is simply a machine room, not dedicated to research, and with no clear plan to deliver services to researchers to address these needs.

Furthermore, planning for the next major data center must be an immediate priority. The MACC is already one-third full, the majority of the space has been committed, and it is expected to be full within two years. There is concern by some units that there may not be sufficient space for the computing and storage components of 2009 faculty start-up packages.

3.8 The College of Engineering-based Center for Advanced Computing (CAC), which is derived from an NSF-funded, U-M-based National Partnership for Applied Computing Infrastructure (NPACI) site, provides an example of how a service facility for research computing could function. The CAC, however, is small, under-resourced, under-marketed, and consequently currently services only a small fraction of "eligible" faculty, primarily in the CoE. CAC, or a campus-wide entity that is the next generation of the CAC, is in acute need of a campus-wide mandate, corresponding vision and resource base, and researcher engagement.

3.9 Future challenges relating to the creation, preservation, use, and sharing of digital data are exploding at an alarming rate. We find that a strategic plan is needed that leverages the broad set of strengths in the campus academic units, university libraries, the U-M Institute for Social Research (ISR) and its Inter-University Consortium for Political and Social Research (IPCSR), IT units, DRDA, the Health System, and the Office of the General Counsel. Regulatory compliance with data-sharing and reporting requirements from sponsoring agencies is an issue demanding immediate attention. Likewise, back-up and security must be addressed campus-wide.

3.10 The U-M has a less than adequate faculty cohort of CI knowledgeable "computational scientists" in certain units, and a large but disconnected faculty cohort in others. A cohort of computational scientists with expertise in parallel computing, large heterogeneous data sets, and algorithm/code development for next-generation computers, including petascale expertise to leverage our participation in the NCSA-based Blue Waters Great Lakes Consortium project, is lacking across all units.

3.11 The current visualization resources on campus are not state-of-the-art and are unable to meet current and projected visualization needs. The facilities are not configured to support cutting edge computational science and engineering, are not integrated into existing computational resources, and do not serve campus units broadly. Computer graphics and scientific visualization experts are needed to support faculty and student needs.

3.12 The numerous blue-ribbon CI panels convened by the National Science Foundation (NSF), other government committees, and private think-tanks provide a more than adequate vision for cyberinfrastructure (annotated references to be found in Appendix I) and its implications for Michigan. All reports note the critical importance of CI for current and future research excellence and competitiveness.

3.13 The committee tried but was unable to acquire research CI expenditure information on individual schools and colleges; thus, an aggregate understanding of current research and instructional CI expenditures at U-M is unavailable. In contrast, administrative IT units that are considered “mission-critical” with appropriately large operating budgets were identified, such as ITCS (Information Technology Central Services; ~$25M/yr), MAIS (Michigan Administrative Information Systems; ~$25M/yr), and MCIT (Medical Center Information Technology; ~$50M/yr). Those three units have an aggregate base budget of $100M, and an aggregate of 1250 base FTEs, not including major initiatives. (Numbers provided by the Office of the Vice Provost for Academic Information.)

3.14 There are several unit-based IT organizations that provide useful CI support to their faculty and students, the most notable being the Computer Aided Engineering Network (CAEN) in the College of Engineering. Other examples are in the Life Sciences Institute (LSI), the School of Public Health (SPH), the Business School, Dental School, the Collaborative Computing and Data Unit (CCDU) of the Center for Computational Medicine and Biology, and others. These unit-based IT service organizations should in general be better supported and coordinated campus-wide.

4.0 Detailed Recommendations

Our recommendations fall within four categories: 4.1) University vision and policies to ensure adequate support of research CI; 4.2) Critical components required for research CI; 4.3) Proposed administrative structure and scope of activities to support research CI, including the coordination of investments and resources; and 4.4) Proposed U-M programs and regional, national, and international partnerships in research CI.

4.1 *University vision and policies to ensure adequate support of research CI*:

4.1.1 The University of Michigan executive officers must recognize the rapidly growing importance of CI to the university research mission, and act boldly to maintain our status as a leading research university in the 21st century by investing in research CI. The University of Michigan must make research CI a “mission-critical” institutional priority, and must commit sufficient ongoing resources comparable in magnitude to other mission-critical IT units at U-M such as ITCS, MAIS and MCIT.

4.1.2 U-M must assess the central and unit-level expenditures for research CI, and leverage these expenditures by a coordinated effort driven by a strategic plan and effective leadership. Certain research CI elements must be centralized for effective leverage and organization, while others should remain decentralized but closely coordinated to facilitate access and accountability.

4.1.3 U-M must assess and coordinate its internal strengths in research CI that exist by virtue of faculty leadership of CI and CI-enabled programs supported by extramural funding, university programs, and school sources. This base of existing programs should be leveraged by a strategic plan and judicious investment to ensure successful competition for future extramural funding opportunities at NSF, NIH, DoE, DoD, and industrial sponsors, including renewal of existing major research programs.

4.1.4 U-M must be proactive to influence and take advantage of proposed and known large r research initiatives from NSF and NIH in cyberinfrastructure, simulation-based engineering and science, data centers, and personalized medicine that depend on CI.

4.1.5 U-M must coordinate its various roles within the Big 10+ research CI group in the Committee on Institutional Cooperation (CIC), and its roles on national and international CI committees and centers to build an overall leadership posture.

4.1.6 To support the emerging University Research Corridor, U-M must become more engaged in the activities of the not-for-profit MERIT Networks, Inc. and coordinate key inter-university research initiatives that depend on CI for success.

4.1.7 U-M must promote the idea of a base allocation of cyberinfrastructure services, just as everyone gets email, printing, and IFS space today. This would give all students, faculty, and staff exposure to the power of modern CI, increasing the rate of adoption and creative use. This base CI would include access to compute cycles and storage for research data. It would also include outfitting classrooms with advanced technologies such as permanently installed automated recording and archiving systems, which would serve a dual purpose of both enriching students' studying experience and also fueling pedagogical research.

4.2 *University-wide mission critical elements of shared research CI:*

The elements of shared research CI combine facilities, computing systems, databases and archives, storage systems, software and algorithmic codes, visualization systems, advanced instrumentation, collaborative tools, and trained personnel – all linked by high-capacity Internet-based communication networks across campus and to appropriate outside research networks and commercial networks. In the following we enumerate recommendations related to these key elements:

4.2.1 Facilities

4.2.1.a A network of well-conditioned machine rooms should be established to house all mid-to-large scale faculty research machines on campus by leveraging economies of scale. These machine rooms should be on separate electrical and communication sub-grids and networked to provide needed redundancies for electrical and networking outages. The research hub facility for the campus should be the Michigan Academic Computing Center (MACC); the research hub facility for the Medical School should be established in the future U-M Health System facility (2010-2011). These hubs should be linked by a networked grid to the many other machine rooms serving research throughout U-M, including central campus, north campus and Arbor Lakes, some of which could serve as needed redundant back-up sites for the hubs.

4.2.1b U-M should initiate the planning process for the design and construction of a MACC 2.0, a next-generation computer facility and data center, with the goal of having an operational facility to complement the MACC in two years, by which time the MACC is projected to reach full capacity.

4.2.1c U-M should publicly declare its intention to develop and exploit opportunities for research on web-based classroom teaching and learning, and take bold steps to outfit all campus classrooms with advanced multimedia recording technologies. The campus should take action to create a research center devoted to evaluating how web-based teaching and learning can be best employed, which techniques work and which do not, and what types of recording systems are best suited for classroom deployment. Within this structure there should be a focused R&D unit directed exclusively to supporting the center's pedagogical work. This will have many additional benefits, including kick-starting the campus OpenCourseWare (OCW) and iTunes U efforts, allow U-M not only to catch up with but surpass other universities such as MIT in lecture archiving and OCW, and provide rich and heretofore unimagined resources and usage data to pedagogical researchers, even enabling classes to be taught in completely new ways.

4.2.1d U-M should establish a campus-wide web lecture recording service to help support the inexpensive web archiving needs of the many research conferences and workshops held on the campus annually. This would permit the broader sharing of research results internal and external to the University, and eliminate the current "hassle-factor" that campus organizers now experience in satisfying mandated archiving requirements for key conferences they convene.

4.2.2 Computing resources

4.2.2a U-M should establish a campus-wide research computing resource available to all faculty, students and staff. This federated computing resource should consist of large compute clusters with both tight and loose coupling, data servers, and more specialized systems such as those with shared memory architecture and new multicore platforms. These systems would be located in the hub facilities described below and several other research grade machine rooms across campus.

4.2.2b U-M should expand and re-purpose the Center for Advanced Computing (CAC), currently run by the College of Engineering. The new center, CAC++, should be developed as a campus-wide center for high performance computing that is a user-focused service provider, serving all participating schools, colleges and units, and overseeing the MACC facility and other federated research CI resources and machine rooms. The CAC++ functions should include:

- Acquisition support, housing and system administration of all high performance research computers and large research data servers, including faculty-acquired research clusters and other platforms, as the CAC currently does for a small subset of (primarily CoE) faculty.

- Provision of open-access compute cycles for research to campus-wide users (students and faculty). These cycles should come both from available nodes on CAC++ supported faculty machines via a "computational kibbutz" style of faculty participation, from a dedicated university terascale platform (see 4.2.2c), and from partnerships with, e.g., IBM, Sun, and other corporations.

- All machine room space administered by the CAC++ would be managed by the CAC++, including the Michigan Academic Computing Center (MACC).

- Development and implementation of a sustainable, cost-neutral plan for long-term hardware/software acquisition, development and support. Create strategic and implementation plans for a heterogeneous set of complementary computing platforms and data servers in addition to the commodity cluster computing machines. These include shared memory and massively parallel systems, highly specialized platforms, etc.

- Assistance, through the use of expert HPC consultants, with code development and implementation for faculty-led research projects. These consultants should be partially supported on faculty research grants and partially supported institutionally.

- Applied research to test research software and hardware to optimize its effectiveness in use.

The committee recommends that, for maximum impact and effectiveness, the CAC++ have a highly visible set of strategically located access points on campus where faculty and students can access system support staff and HPC consultants. Dedicated staff within these access points should be assigned to support individual units, and embedded within the unit they serve.

4.2.2c U-M must work with a vendor to establish a partnership for implementing a large terascale supercomputing system at U-M (> 10 Tflops) to nucleate and support high performance computing research activities, and to ensure U-M competitiveness. This system would be considered a Tier 3 system in the NSF HPC hierarchy. Such a system is the minimum bar for credibility in the accepted research computing hierarchy recommended by the NSF. Funding for the first of such systems must be provided for by the institution. This system would complement NSF Tier-2 Distributed Terascale Facility (DTF) and NSF Tier-1 Petascale resources (see section 4.2.2d).

4.2.2d U-M should assist researchers in accessing a coordinated set of national computational resources, providing access to Tier 1 and 2 resources. This coordination function would work with campus units and researchers to access computing cycles from national partners such as the NSF Distributed Terascale Facility (DTF), which includes the San Diego Supercomputer Center (SDSC), Pittsburgh Supercomputing Center (PSC), and the Texas Advanced Computing Center (TACC); NIH Resources such as the National Biomedical Computing Centers, the Biomedical Informatics Research Network (BIRN), the CTSA Informatics Consortium, the cancer BioGrid (caBIG); DoE resources, including national laboratory sites and NERSC; and DoD computing facilities at the services laboratories and other sites. In addition, the U-M should be aggressive in establishing partnerships with IBM, Sun, and other computer vendors to facilitate researcher access to compute cycles.

4.2.3 <u>Data Resources and Assurances</u>

4.2.3a U-M should initiate a campus-wide planning activity to determine the well-enunciated policies leading to physically and digitally secure IT infrastructure relating to the emerging issues surrounding the security and privacy of research data, especially those data, such as DNA sequences, which might contain unique personal identifiers.

OVPR, together with DRDA, should scrutinize, establish, and implement U-M research datasharing policies, in cooperation with the Schools, Colleges, Institutes, and Centers, that are in line with Federally mandated policies, rules and stipulations set down by government research sponsors.

4.2.3b U-M should coordinate operational aspects of establishing an ongoing framework of digital libraries of research data at U-M. The University Library, ISR and ICPSR, and the U-M Health System should be key members of this planning group. Each school and college should be directed by the Office of the Provost to participate in this effort. Elements such as standardized digital research laboratory notebooks and digital Laboratory Information Management Systems (LIMS) should be actively considered for campus-wide adoption by U-M researchers. This should complement and build upon ongoing U-M "Deep Blue" archiving activities.

4.2.3c U-M data stewardship activities should leverage IT production systems at U-M such as MAIS and MCIT, and ISR resources, which have state-of-the-art security and privacy systems.

4.2.3d U-M should take concrete steps to assure that the Google Library Scanning Project will benefit the research (and instructional) missions by giving special access to our holdings and creating programs to lead in its early use.

4.2.4 Software Licensing and Availability

4.2.4a U-M should establish a centralized scientific software clearing house that would coordinate purchases of proprietary research software for U-M researchers, maintain a digital library of useful and current versions of open source software, and also make available U-M-derived software modules and codes to U-M researchers. Open-access promotion nationally or globally of U-M-developed software should be a priority. This office would also interface with the Technology Transfer Office (TTO) in OVPR relating to Intellectual Property (IP) issues for Michigan-created software.

4.2.5 Networking and Middleware Activities

4.2.5a IT-Com should bring multiple 10 gigabit/sec optical connections into selected 'high-impact' collaborative zones in several schools, colleges, institutes, centers and laboratories across campus through an optical test-bed that allows the switching of connections for both internal and external connections. This would amount to a build-out of a "control-plane" infrastructure that facilitates wave and sub-wave allocation of "light paths" within campus and to remote locations, supporting the bandwidth and quality of service demands of advanced research (and instructional) applications. Existing research activities like TeraPaths, VINCI and UltraLIght from the high energy physics community can be leveraged to support the prototyping and build-out of such an infrastructure. Connectivity with researchers at other institutions is critical for multi-institutional large-scale research.

4.2.5b OVPR should provide opportunities for funding to use this network (described in 4.2.6a) for appropriate team-science collaborative activities and high definition (HD) video teleconferencing and research data sharing systems, building from systems pioneered by the School of Information (SI), College of Engineering, and Medical School. An example is the Virtual Space Interaction Test-bed Project (VISIT), a collaboration between the School of Information and the Office of the Provost to build and deploy Opti-Portals for ultra-resolution collaboration across campus. It focuses on HD videoconferencing, multi-megapixel visualization, advanced networking, and sensor networks.

4.2.5c U-M should review the current Michigan Lambda Rail (MiLR) connection pricing and service model to make it more readily accessible to faculty and researchers. Strong incentives should be offered by OVPR for National, Regional and State of Michigan Team-Science Research (or Instructional) projects. U-M should identify and deploy research applications (bulk data transfer, shared facilities, and advanced collaboration environments) that will take advantage of those network connections. The U-M School of Information has recently equipped three endpoints on MiLR (two at U-M and one at WSU) to host joint live events between U-M and WSU, with MSU to be added in 2008.

4.2.5d U-M should establish an architecture and implement an infrastructure for virtual organizations that addresses identity, authentication, authorization, and resource discovery and access in a federated, multi-institutional environment. The goal is for faculty and researchers to use their local identities to participate in national and international collaborations. This would include using collaboration tools, sharing data, executing codes remotely, and much more as a natural extension of the local environment. This would enable distributed and federated research (and instructional) virtual grids.

4.2.5e U-M should deploy software to "scavenge" computing cycles from student computing laboratories and more generally from idle desktop machines as CAEN currently does with CAEN lab computers. This "condor flock" technology was demonstrated by the Michigan Grid Research and Development (M-GRID) Center several years ago, and was able to supply significant computing cycles to the Biophysics Research Division (BRD) scavenged from the LS&A student computing laboratory in Angell Hall.

4.2.6 Instrumentation and Visualization

4.2.6a Remote Cyber-enabled Instrumentation: U-M should establish partnerships with remote sites (supercomputer centers, telescopes, microscopes, visualization facilities, and collaboration locations) to provide ongoing access to resources of interest to U-M faculty researchers and trainees that cannot be provided locally.

4.2.6b Expand visualization facilities and integrate with CAC++: U-M should expand its existing visualization facilities in the Duderstadt Center and create a small set of grid-enabled satellite nodes at Central Campus, UMMS, and the Dearborn and Flint Campuses. Expert computer graphics and scientific visualization staff should be maintained as consultants, partially supported on faculty research grants, partally supported institutionally. In order to adequately sustain such a campus CI resource, several new faculty and staff must be recruited with interested schools and colleges (e.g. CoE, Health Sciences Schools, Art and Design, Architecture and Urban Planning, etc.). Visualization is a key component of realizing the promise of research CI, as the extreme complexity of modern datasets requires new mathematical and statistical models, algorithms, software, screens and a creative and collaborative environment.

4.2.7 Cyberinfrastructure Personnel

4.2.7a Specialized Staff: The committee recommends that three types of staff be recruited:

> 4.2.7a.i Systems administration staff experienced in research computing, database, storage, and networking systems are needed to administer all R-CI resources maintained by the CAC++.
>
> 4.2.7a.ii Research staff expert in high performance computing, especially massively parallel computing and next generation architectures, as well as big data, visualization, and software development. This HPC consultant service is critical to support leading computational-based research. We recommend hiring the minimum number sufficient to cover the major research computing thrusts at U-M (e.g., molecular and particle-based simulations, quantum mechanics computations, field-based simulations, optimization, informatics, etc.), each of which cuts across multiple disciplines and departments (e.g., energy and sustainability, materials, nanoscience, medicine, biology, etc.). At steady state, these HPC consultants should be partially supported by faculty research groups through research grants and other sources.
>
> The committee further recommends that U-M acquire (in part by growing from within) petascale experts who would serve as an interface to the NSF Petascale Project at NCSA and work closely with research scientists at NCSA to facilitate faculty and student access to petascale resources.
>
> 4.2.7a.iii Research data specialists with expertise in database systems creation, metadata tagging, ontology creation, data integration, standards implementation, data curation and archiving are required.

After consideration of the minimal staffing levels for functional HPC facilities in the Big10+ and comparable Carnegie R1 universities, the committee has determined that the CAC++ needs to hire 10 FTEs by 2009 towards a total of 30 FTEs by 2013. This number corresponds to our finding of a minimal staffing level to achieve critical mass in a campus-wide HPC and R-CI facility.

4.2.7b Partnerships with existing IT units and personnel: The CAC++ and other R-CI activities at U-M should take advantage of an active partnership with IT production service organizations at U-M. These units should become partners for many operational aspects of R-CI.

4.2.7c Faculty: U-M should establish at least 15 partially funded or endowed Instructional and Research Track lines for faculty in emerging CI areas. Petascale computing faculty are needed to build our strengths in parallel algorithms, operating systems and languages for multicore architectures, complemented by computational science applications-oriented faculty. Data science faculty are needed to build our strengths in areas of database, preservation, metadata, and web services for data.

Coordination across departments and/or units should be encouraged to recruit "cyber" faculty whose cutting edge, high performance computing-based and data science-based research spans computer science and/or information science, and one or more application domains such as energy or nanoscience. Hiring must be coordinated across the ongoing recruitment drives in the various units, the Office of the Provost, and the Office of the President. Such hires are needed to move U-M ahead of our peer institutions in the area of cyber- enabled discovery and innovation and position us for leadership in this emerging, cross-cutting area. Also, U-M should organize an effort around faculty recruitment and retention that leverages the IT-related dollars that typically accompany faculty packages toward access to cyberinfrastructure services.

4.3 *Administrative structure for the support of research CI:*

The above recommendations are aimed at positioning U-M as a 21st century leader in CI and its use in discovery and innovation. Implementation of these recommendations will allow Michigan to become a national leader in research CI and CI-driven research within five years. To attain this ambitious and necessary goal, the committee recommends the following administrative structure:

4.3.1 U-M should establish an Office of Research Cyberinfrastructure (ORCI) within the Office of the Vice President for Research (OVPR). The recommended creation of a research-centric CI office is based on the committee's evaluation of practices at other Tier 1 research insitutions, and recognition of the need for a tighter coordination between CI and the unique requirements of researchers.

We suggest that the director of ORCI should be a faculty member with a 50% appointment in OVPR, and should hold the title of Associate VPR of R-CI, and report directly to the VPR. The ORCI would have the mandate and budget to lead the creation of a federated CI system for research that spans the University of Michigan. The ORCI would provide OVPR with clear institutional planning, budget recommendations, and proposed integration of existing and future large-scale, interdisciplinary campus CI activities. The ORCI will create institutional priorities, and be responsive to the opportunities provided by state and federal funding agencies and programs. Creation and implementation of the vision described in section 4.1 and oversight of all critical research CI components described in section 4.2 would be the responsibility of this office. The ORCI should be the primary internal and external point-of-contact and coordination for all research-CI inquiries and new large-scale university activities.

4.3.2 U-M should establish a well-supported and highly visible Cyberinfrastructure Operations Center (CIOC) within the Office of the Vice Provost for Academic Information to coordinate efforts among ORCI and existing IT organizations on campus, including ITCS, CAEN, and MSIS, to provision IT services where appropriate. CIOC will also coordinate with the Libraries and ORCI in support of research CI.

4.3.3 Funding for research CI should comprise a centrally-funded operating base within ORCI, with additional contributions from the participating units across U-M, endowment(s), and a percentage of IDC from team-science-led, extramurally-funded initiatives.

4.3.4 The ORCI Director should be responsible for organizing CI-driven, large-scale faculty research efforts, and building and sustaining partnerships across campus and with external organizations such as Merit Networks, Inc., the Great Lakes Consortium for Petascale Computation (GLCPC), and NCSA. The ORCI Director will spearhead large proposal efforts to secure CI resources from federal agencies and other sources, and assist U-M researchers in obtaining access to external computational and data resources (such as NSF supercomputing centers, NERSC, Teragrid, and NIH initiatives).

4.3.5 The CAC++ administration and operations should fall under the auspices of the ORCI. The director of CAC++ should be a full-time, senior technical staff member, hold the position of Associate Director of ORCI, and report to the Associate VPR of R-CI. The CAC++ director should chair an Operations Committee comprised of IT leads from all participating schools and colleges. The Operations Committee should be responsible for implementing all policies crafted by the ORCI Policy Committee (see 4.3.6) that relate to research computing facilities and personnel. Oversight and guidance of the CAC++ would be provided by the ORCI Policy Committee. Responsibility for the MACC, which serves as the primary facility of the CAC++, would thus be moved to OVPR under the auspices of the ORCI/CAC++. The CAC++ Operations Committee would be responsible for implementing MACC policies set forth by the ORCI Policy Committee.

4.3.6 The Associate VPR of R-CI should chair the ORCI Policy Committee, comprised of, the ORCI Associate Directors, leading research CI faculty from participating units, and IT staff leads from the major units. The Policy Committee will be responsible for crafting the overarching governing policies for research CI support and administration across U-M, including policies for sharing and unit support of R-CI facilities, including CAC++. In particular, the ORCI must lead in the planning and execution of the research computing and data facets of the Michigan Academic Computing Center (MACC) and the federated grid of distributed research-CI facilities across campus.

4.3.7 An Office for Research Data Stewardship (ORDS) should be established within ORCI, and led by an associate director of ORCI. ORDS would be responsible for all research data resources and assurances described in sections 4.2.3a-c.

4.3.8 The ORCI should create a position for a Business Development staff person who would work closely with U-M and unit development offices to assist the Associate VPR for R-CI in forging partnerships with industry (e.g. IBM, Intel, Apple, Microsoft, Google, Amazon, etc.) and securing donations from alumni and others to help sustain R-CI.

4.3.9 There must be established metrics to guide and evaluate the U-M investment in research cyberinfrastructure. These metrics would include those for achievement, productivity, and budgets of existing U-M IT infrastructure units, and must be related to data from peer universities. ORCI should direct this effort. Some data have already been collected or have been requested by the R-CIC.

4.3.10 An active ORCI External Advisory Committee should be established. This committee would provide external guidance to OVPR, the Office of the Vice Provost for Academic Information, the University Library, and ORCI. Recommendations for potential founding ORCI-AC members include alumni or current or past faculty of U-M who have considerable expertise in research CI: Mark D. Adams (Case Western Reserve University), Frances Allen (IBM), Dan Atkins (formerly NSF Office of CI, on loan from U-M), John Seely Brown (Xerox Parc), Ralph Cicerone (NAS), Francis Collins (NIH, IOM), Jim Duderstadt (U-M, NAE, NSF Advisory Committee on CI, Spellings Commission), John Evans (Evans Telecommunications Co.), Larry Page (Google), Doug van Houweling (Internet2), Tim Killeen (NCAR), Charles Vest (NAE), and Robert Weisbach (Woodrow Wilson Institute).

4.4 *Proposed CI programs and partnerships:*

Presented below are several critical CI programs and partnerships that must be addressed as part of a comprehensive CI plan for the university.

4.4.1 The cross-campus Rackham Certificate Program in Scientific Computing, currently based in CoE, should be revitalized. This program was one of the first of its kind in the nation, and served as the model for the DOE's highly respected Krell Computational Sciences Graduate Fellowship program. The program should provide critical training of computational science and engineering students in simulation and informatics to complement their primary departmental degree. The program should reach out to students across the U-M and its requirements should be modernized. The program should be coordinated closely with the Great Lakes Consortium for Petascale Computation's (GLCPC) Virtual School for Computational Science and Engineering (see 4.4.2). This coordination will leverage new activities to be developed under the Virtual School, and provide leadership to CIC institutions in certificate programs in scientific computation.

4.4.2 Great Lakes Consortium for Petascale Computation: The ORCI will serve as the point of contact for U-M's involvement in the GLCPC within the Committee on Institutional Cooperation. NCSA is the home of the GLCPC. The plan will identify incentives to encourage participation and collaboration from schools, colleges, and units as well as existing, newly established, and proposed research centers across campus. The ORCI will be responsible for the formulation and communication of this plan. The ORCI and CAC++ will be the U-M node in the GLCPC Virtual School of Computational Science and Engineering.

4.4.3 Grid Middleware and Global Research Computing Initiatives: The Open Science Grid (OSG--www.opensciencegrid.org) has emerged as a positive new force in research CI; it is analogous to the eScience project in the UK, and the Enabling Grid for E-sciencE (EGEE). U-M must leverage its membership in the OSG by investing in this initiative, building from the strengths of strong projects such as the DoE-based ATLAS physics project and the NIH-based NCIBI. U-M must reassess its relationship with Internet2, to determine if U-M is still deriving sufficient value, compared with the potential of redirecting these investments into partnerships with other initiatives such as the National Lambda Rail (NLR).

4.4.4 U-M CI Pilot Grants Program. ORCI should initiate a pilot grant opportunity program to stimulate CI research and development activities across campus, with special emphasis on the arts, humanities and the social and behavioral sciences. These pilots will be strategically oriented toward leveraging important computing intensive cross-campus centers such as the U-M Center for Computational Medicine and Biology (CCMB), the Michigan Nanotechnology Institute for Medicine and Biology (M-NIMBS), the Space Weather Initiative, the US –ATLAS projects, and the proposed U-M Materials Research Science and Engineering Center (MRSEC) for Cyber-enabled Materials Discovery, Design and Innovation.

4.4.5 Green Energy CI Initiative: ORCI should work with the U-M Graham Environmental Sustainability Institute, the Michigan Memorial Phoenix Energy Institute, the School of Natural Resources and the Environment (SNRE) and the Vice President for facilities to establish a green computing initiative. This has the promise to save the University significant electricity charges, likely in excess of $1M per year, while significantly reducing greenhouse emissions. ORCI should work with the Office of the Vice Provost for Academic Information to join the International Green Computing Initiative (see www.climatesaverscomputing.org).

4.4.6 National Presence at Supercomputing Meetings (and other high-visibility venues): ORCI should set the goal of having a University of Michigan booth at SC09 in Portland, Oregon, and future follow-on (and related Tier-1) meetings. Such a goal could be a forcing function for us to bring our disparate efforts together into a coherent story and allow us to identify the initiatives that would make up a compelling showcase.

5.0 References

Athey, B.D. et al. 2007. NIH National Center for Integrative Biomedical Informatics (http://portal.ncibi.org/portal).

Atkins, D.M., National Science Foundation, 2006. “NSF’s Cyberinfrastructure Vision for 21st Century Discovery.” Version 7.1. http://nsf.gov.

Atkins, D.M. 2007. “Where the Rubber Hits the Road” Discussion on the NSF Office of Cyberinfrastructure (OCI). 2007. NCSA *Access*, *Vol. 20*, no. 1.

http://access.ncsa.uiuc.edu/Stories/Atkins/.

Atkins, D.M. “Creating a National Cyberinfrastructure for Science and Engineering (and nurturing it and benefiting from it)”. Talk given July 13, 2007 at the Great Lakes Computing Consortium meeting, Northwestern University, Evanston, Illinois.

Benioff, Marc. R. and Edward D. Lazowska, Co-chairs. “Computational Science: Ensuring America’s Competitiveness.” 2005. President’s Information Technology Advisory Committee (PITAC). http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf.

Berman, F. and H. Brady. 2005. “Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences.” http://vis.sdsc.edu/sbe/reports/SBE-CISE-FINAL.pdf.

Duderstadt, J.J., William A. Wulf, Robert Zemsky, and Thomas Arrison. 2005. “The IT Forum: Preparing for the Revolution Redux.”

http://milproj.ummu.umich.edu/publications/it_forum_summary/download/IT%20Forum%20Summary.pdf.

Duderstadt, J.J. "A Lesson from the Past", July 9, 2006. White paper submitted to Brian Athey, Sharon Glotzer, Tom Finholt, and John King.

Glotzer, S.C. et al. "International Assessment of Research and Development in Simulation-based Engineering and Science: Report of a US Baseline Workshop," Nov. 1-2, 2007. http://www.wtec.org/sbes

Gutmann et al. 2007. ICPSR Inter-University Consortium for Political and Social Research http://www.icpsr.umich.edu/.

Hacker, T.J., Wheeler, B.C. 2007. "Making Research Cyberinfrastructure a Strategic Choice, growing demands for research computing capabilities call for partnerships to build a centralized research cyberinfrastructure." *Educause Quarterly, no. 1*, 21-29.

Neal, H.A. and S. McKee, Michigan US Atlas Physics Project, http://ganesh.physics.lsa.umich.edu/.

Neuroth, H., M. Kerzel, and W. Gentzsch, Editors. 2007. "German Grid Initiative D-Grid." http://www.d-grid.de/index.php?id=1&L=1.

Oden, T, et al. "NSF Blue Ribbon Panel Report on Simulation-based Engineering and Science," National Science Foundation, May 2006. http://www.nsf.gov/pubs/reports/sbes_final_report.pdf.

Uhlir, P. F. and J. M. Esanau, Rapporteurs. 2006. "Strategies for Preservation of and Open Access to Scientific Data in China", Summary of a Workshop. National Research Council of the National Academies. 131 pp. http://www.nationalacademies.org/gateway/pga/3634.html.

Vest, C.M., "Open Content and the Emerging Global Meta-University, Educause Review vol. 41, no. 3 (May/June 2006): 18-30.

http://www.educause.edu/apps/er/erm06/erm0630.asp.

Weil, Peter and Jeanne W. Rosse. "IT Governance: How top performers manage IT decision rights for superior results." 2004. Harvard Business School Press, Boston, MA. ISBN 1-59139-253-5.

Welshons, Marlo, Editor. "Our Cultural Commonwealth: The report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences." http://www.acls.org/cyberinfrastructure/.

Association for Research Libraries (ARL): "To stand the test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering." A report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe. September 26-27, 2006. http://www.arl.org/info/events/digdatarpt.pdf.

For more useful ARL whitepapers see: http://www.arl.org/pp/access/nsfworkshop.shtml.

NSF Petascale Announcement mentioning great Lakes Computing Consortium. http://www.nsf.gov/news/news_summ.jsp?cntn_id=109850.