OFFICE OF THE PROVOST

AND VICE CHANCELLOR FOR RESEARCH

WORKING GROUP IN COMPUTATION, DATA, AND INFORMATION

FINAL REPORT

January 6, 2015

TABLE OF CONTENTS

I. COMMITTEE CHARGE (DECEMBER 2013)

I am writing to invite you to serve on the Interdisciplinary Working Group for Computation, Data and Information, which will be chaired by Prof. Bruce Hajek. This Working Group, jointly sponsored by the Office of the Provost and the Office of the Vice Chancellor for Research, will advise campus administration on communications and other activities to support the growth, strength, and visibility of the campus's interdisciplinary research in these areas. The goals of this Group will include:

- Advising campus administration on how best to support both scholarship in these areas and partnerships between researchers in these areas and scholars in other disciplines
- Building on the process of Visioning Future Excellence, connecting individuals and groups of researchers to enable creative new scholarship, as well as assisting researchers in connecting emerging capabilities with a range of scholarly disciplines
- Working with OVCR staff to create strong coherent external communications that present Illinois as a world leader in computation, data and information
- Preparing the campus for more effective pursuit of large funding opportunities in these areas by identifying them early and suggesting possible proposal teams
- Identifying the multiple units, centers, and research programs on campus that support research in these areas and making recommendations for how to promote possible synergies

I look forward to working with the Computation, Data and Information Working Group to establish milestones and timelines that will help guide its activities, while also allowing for flexibility in response to new developments. It is my hope that by the end of 2014, this Working Group will be broadly recognized for providing greatly improved communication on campus for activities in its area and for broadly serving faculty research programs. I will be particularly interested in the Working Group's recommendations for how we can institutionally provide better support so that our researchers can maximize their potential for exciting interdisciplinary, collaborative work that will both distinguish the University of Illinois and contribute to "finding solutions to the grand challenges of the century," as Chancellor Wise and Provost Adesida charged in the letter accompanying the 2013-2016 Strategic Plan.

Best regards,
Peter Schiffer
Vice Chancellor for Research
Professor of Physics
University of Illinois at Urbana-Champaign

II. COMMITTEE MEMBERSHIP

- Bruce Hajek (ENGINEERING, ECE), Chair
- Doug Simpson (LAS, Statistics), Vice-Chair
- Scott Althaus (LAS, Political Science)
- Narayana Aluru (MechSci and Engineering)
- Minh Do (ENGINEERING, ECE)
- Miles Efron (GSLIS)
- Jeff Erickson (ENGINEERING, Computer Science)
- Carl Gunter (ENGINEERING, Computer Science)
- Jiawei Han (ENGINEERING, Computer Science)
- So Hirata (LAS, Chemistry)
- Matthew Hudson (ACES, Crop Sciences)
- Bill Kramer (NCSA)
- Zhi-Pei Liang (ECE, Beckman)
- Melanie Loots (OVCR)
- Luke Olson (ENGINEERING, Computer Science)
- Carole Palmer (GSLIS)
- Rob Pennington (NCSA)
- Luis Rodriguez (ACES and ENGINEERING, Agricultural and Biological Engineering)
- Michael Shaw (BUSINESS)
- Bill Spencer (Engineering, Civil and Environmental)
- Lisa Stubbs (LAS, Cellular and Developmental Biology)
- Ted Underwood (LAS, English)
- Chengxiang Zhai (ENGINEERING, Computer Science)

The Chair and Vice-Chair thank the members of this committee for their contributions and service, and to VCR staff members Andy Blacker and Melissa Edwards for their logistical support.

III. COMMITTEE ACTIVITIES

Having been chartered in December 2013, the Committee met nine times between December 16, 2013 and December 3, 2014 with guests as noted:

- 12-16-2013, 1-2:30 pm
- 2-3-2014, 10:30 am – noon
- 3-6-2014, 3-4:30 pm
- 4-2-2014, 3:30-5 pm
    - o Guests: Laura Frerichs (Research Park), Michael Welge (NCSA, Research Park)
- 5-7-2014, 10:30 am – noon
    - o Guests: Jeff Binder (ARI), Ed Seidel (NCSA), Gabrielle Allen (NCSA)
- 9-3-2014, 3:30-4:30 pm
- 10-6-2014, 3:30-4:30 pm
- 11-4-2014, 3:30-4:30 pm
- 12-3-2014, 3:00-4:00 pm

Committee members shared information through a Wiki site hosted by CITES. The committee compiled and examined listings of (A) Computing resources at Illinois (i.e. institutes, centers, laboratories, and groups on campus with missions aligned with developments in computation, data, and information; (B) Computing events and symposia on campus; and (C) Peer institutes and consortiums. See the appendices. The committee met with leaders of the research park, the Advanced Research Institute (recently launched within the research park), and NCSA.

At most of its meetings, the committee deliberated about how to grow and strengthen connections within the university (connecting researchers to each other, to campus resources and services, and to outside funding organizations) and externally communicating Illinois expertise, strengths and successes in computation, data, and information.

IV.   FINDINGS AND RECOMMENDATIONS

Findings:

As detailed in Appendices A and B, there are many exciting and innovative research activities taking place on campus in the broad area of computation, data, and information: from NCSA and Blue Waters, to Grainger Engineering Breakthroughs Initiative (GEBI) hires in big data, from research park industrial growth, to the Applied Research Institute. Still, due to the recent current explosion in availability of data, and tools to handle it, we anticipate much additional interest from industry, government funding agencies, and private foundations in projects involving computation, data, and information.

Certainly our peer universities are aggressively responding to recent developments and opportunities in big data, through the creation of new programs, cluster hires, institutes, etc., as detailed in Appendix C. Many of the opportunities for future collaboration in this space involve a broad spectrum of research, from the humanities and fine and applied arts, to science and engineering.   In the long term, campus researchers should continue to work across disciplinary boundaries to engage new opportunities. In the near term, an anticipated program by the National Science Foundation for regional big data hubs is an impetus for active planning.

Some unique features of the UIUC campus in the areas of computation, data, and information include the major research center - NCSA, very highly ranked academic departments in all related areas, several other major institutes and laboratories, a research park with rapidly growing interest in data analytics, the recently launched Applied Research Laboratory in the research park, the GEBI focus on big data, and two large, modern computing and data facilities (Advanced Computation Building and the National Petascale Computing Facility).   Important research activities within computation, data, and information are distributed across multiple Colleges and Institutes on the campus, with computer science and engineering in the College of Engineering, the Department of Statistics in the College of Liberal Arts and Sciences, the Institute for Genomic Biology (IGB), the Beckman Institute, and significant activities within the Graduate School of Library and Information Sciences (GSLIS), the College of Business, the College of Agriculture, Consumer and Environmental Sciences (ACES), and others colleges.

NCSA is coming to play an increasing role on campus to help make connections across campus. A challenge is to bridge not only college boundaries, but also traditional academic and scholarly boundaries to embrace and support the full range of opportunities to incorporate information and data oriented thinking.

Recommendations:

Based on our investigations, and with the above findings in mind, the committee recommends the campus continue its strong support of research in the area of computation, data, and information.   For example, support of institutes, laboratories, the research board, and centers, as well as matching funds for faculty startup packages and seed funding to jump start collaborative projects are critically important. Particular projects that may be of interest in the near future could involve groups working on domain-specific data sets to create large-scale data repositories. The committee endorses the campus effort to institute a system for faculty members and other researchers on campus to link their research profiles. To complement these important activities, the Committee offers the following three recommendations:

Recommendation 1. OVCR should ask/encourage NCSA and the Coordinated Science Laboratory to jointly organize a two day workshop in the next year (ideally in Spring 2015). The two day event would help lay the groundwork for a campus proposal or proposals for a regional data hub.   Representatives and participants from other departments and organizations around campus should be invited to participate, including

- All academic departments (Astronomy, CS, ECE, Stats, and any others interested)

- Beckman Institute, Institute for Genomic Biology (IGB), Micro and Nano Technology Lab, Computational Science and Engineering (CSE), GSLIS/ITI, Research Park, ARI,

- Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) ,

- The Cline Center for Democracy,

- eDream - the Illinois Emerging Digital Research and Education in Arts Media Institute

- Center for IT and e-Business Management

It is recommended that each organization be limited to two presentations. One presentation could be an overview style talk that describes some of the work being done in the organization and capabilities particularly relevant to a possible regional data hub. The other presentation could be a technical talk by a faculty member or other investigator in the organization that exemplifies work being done in the unit. This could be thought of as a dry run for a site visit by a panel of outside experts.

Recommendation 2. It is recommended that the campus launch one or more faculty hiring clusters in the area of computation, data, and information. This could complement the GEBI program in Engineering aimed, in part, at senior hires in the area of big data. The hires could be in the Arts, Social Sciences, or Humanities, or in Engineering at the junior level, to complement the GEBI hires. This committee recognizes that faculty and relevant units across campus should be the ones to settle on appropriate areas for cluster hires, but our committee came up with some suggestions that might be considered, including:

- Information and Technology in Arts, Social Sciences, and Humanities
- Social Sciences and Humanities Research in Digital Rights and Privacy
- Computational Health Science and Engineering
- Personal Learning Environments
- World Population, Religion, and Education: Modeling, Prediction, and Policy
- World Food Security: Biological, Agricultural, and Social Challenges
- World Water Resources

All the above suggestions are for cluster hires strongly intertwined with computation, data, and information. For example, the cluster hires could possibly foster groups working on domain-specific data to create large-scale data repositories. This is important because as the data become bigger and bigger, the research activities and funding opportunities in the future will likely be centered around where the repositories are, because of the difficulty of moving very large data and, in some cases, privacy issues.

Recommendation 3. It is recommended that OVCR complete the development and launch of a light-weight web portal linking to the many institutes, centers, departments, and groups on campus with activities in the computation, data, and information space. It is recommended the OVCR form a standing committee of about six individuals from across campus with interest in the communications project, to meet with OVCR staff once or twice per semester, to oversee the website. The website would help researchers across campus make connections, and help external stake holders identify campus resources, despite the sheer number and diversity of activities on campus. The website would also serve to motivate the constituent centers and groups to keep their own pages up to date, particularly if the OVCR staff were to assess the freshness of the information and identify stall/broken links as part of the oversight. It should be possible to retire and archive old or static web sites so there is still overall access, but the information being presented first in the portal should be the most active

Appendix A: Computing Resources at Illinois

Beckman Institute

An interdisciplinary research institute founded on the premise that reducing the barriers between traditional scientific and technological disciplines can yield research advances that more conventional approaches cannot. It is devoted to interdisciplinary research in the physical sciences, computation, engineering, biology, behavior, and cognition, focused around four research themes:

Beckman Institute research is focused around four research themes:

*   Biological Intelligence
*   Human-Computer Intelligent Interaction
*   Integrative Imaging
*   Molecular and Electronic Nanostructures

The Beckman Institute is also home to three strategic initiatives that seek to unify campus activities in their respective areas:

*   Health: Attitudes, Biology, Information, Technology, Society
*   Strategic Initiative on Imaging
*   Social Dimensions of Environmental Policy

More than 1,500 researchers from 43 departments are affiliated with the Beckman Institute.

Blue Waters Sustained Petascale Computing and
Additional Blue Waters information

The Blue Waters, a project led by NCSA, is a supercomputer that is the most powerful system located at any university campus with an impressive 14.1 Petaflops/s (PF) peak performance, 1.66 Petabytes (PB) of memory, 26.4 PB of usable, online disk storage and a 300+ petabyte nearline/archive storage system. The balanced architecture of the system makes it one of the world's most powerful and capable supercomputers available for compute-intensive and data-intensive research.

Over 700 scientists and engineers from across the country and around the would are using Blue Waters to advance our knowledge of disease, climate, cosmology, solar storms, severe weather events, earthquakes, medicine, materials, and much more. Over 250 researchers from Illinois are part of the Blue Waters user community, participating in almost ½ the project allocations of Blue Waters. The Blue Waters teams focused on providing a very reliable and performant computational and data analysis resource, genomics, industrial processes, etc. to almost 200 science and engineering teams that have access to Blue Waters. Many of these S&E teams are

performing simulations at scales or with degrees of accuracy that has not been possible without Blue Waters. Blue Waters makes approximately 187 million node-hours (about 6 billion integer core-hours) of computing resources available annually to the national research community. There is a national NSF run allocation process for at least 80% of the Blue Waters resources, and an Illinois run process for approximately 7%. In order to maximize the productivity of the teams running on Blue Waters, NCSA experts are providing advanced user support with code optimization and scaling, optimizing data transfer, advanced visualization, etc.

Coordinated Science Laboratory (CSL)

An interdisciplinary research laboratory with a history of more than 60 years of innovation, CSL focuses on information technology at the crossroads of computing, control, and communications, exploring issues such as defense, medicine, environmental sciences, robotics, life enhancement for the disabled and aeronautics. Researchers from more than a dozen disciplines have looked at:

- Circuits,
- Communications,
- Computational and physical electronics,
- Decision and control,
- Reliable and high performance computing,
- Remote sensing and space science,
- Signal, image and speech processing,
- Thin film electronics.

CSL has created centers and institutes in several key areas, including integrated microsystems, wireless systems, communication technologies and trustworthy supercomputing. Many of the centers have strong ties with industry and government partners. CSL's interdisciplinary strength has led to the formation of several institutes including the Advanced Digital Sciences Center in Singapore, the Information Trust Institute and the Parallel Computing Institute.

Institute for Genomic Biology (IGB)

An interdisciplinary institute established to advance life science research and stimulate bio-economic development in the state of Illinois. Dedicated to transformative research with program areas in systems biology, biomedical research, cellular and metabolic engineering and genome technology, the IGB pioneers advances in the life sciences. Genomic biology is one of the key "big data" disciplines, and the IGB CNRG group maintains biology-dedicated high-performance computer systems with thousands of processors and petabytes of storage.

Micro and Nanotechnology Laboratory (MNTL)

MNTL is a multidisciplinary research facility in the College of Engineering that supports research in photonics, microelectronics, nanotechnology, and biotechnology. The laboratory facilitates research activities in optoelectronics and photonic systems, microelectronics for wireless communications, microelectromechanical systems, and nanobiosystems. MNTL is **one of the**

nations largest and most sophisticated university-based facilities for semiconductor, nanotechnology, and biotechnology research. It currently houses the DARPA-funded Center for Bio-Optoelectronic Sensors and Systems.Research in areas such as bioengineering generate enormous amounts of data to be stored and analyzed.

National Center for Supercomputing Applications (NCSA)

The National Center for Supercomputing Applications (NCSA) provides computing, data, networking, and visualization resources and services that help scientists and engineers across the country better understand our world. NCSA is an interdisciplinary hub and is engaged in research and education collaborations with colleagues and students across the campus of the University of Illinois at Urbana-Champaign. The center focuses on BIG COMPUTING, BIG DATA, and BIG RESEARCH

Established in 1986 as one of the original sites of the National Science Foundation's Supercomputer Centers Program, NCSA is supported by the state of Illinois, the University of Illinois, the National Science Foundation, and grants from other federal agencies.

The center focuses on:

- Data and Information Science & Technology: Aggregating distributed, heterogeneous datasets at all scales and analyzing them to obtain reproducible results with scalable, maintainable software pipelines on optimized computational data platforms in highly efficient, secure collaborative environments.
- @Scale Science & Technologies: Deploying, supporting and improving computational, analytical, storage and communication technologies and services at capabilities and complexities of notable scale.
- Collaborative eScience: Creating and managing distributed and integrated environments that support the inclusion of a diverse range of advanced computing resources and services to accelerate the discovery and application of new knowledge.
- Industry: Helping industry partners take advantage of advanced computing and digital ecosystems to enhance their competitiveness.

Research Park

The Research Park provides an environment where technology-based businesses can work with faculty and students to take advantage of opportunities for collaborative research and easy access to University labs, equipment and services. The Research Park is now home to more than 90 companies and growing, employing more than 1,400 people in high-technology careers. At any given time more than 400 student interns are working in these companies gaining valuable work experience while making real contributions to internal corporate R&D and product development programs. The Research Park is also home to 50+ startup companies that are commercializing technology. EnterpriseWorks, the Research Park's 43,000-square-foot business incubator for early-stage tech firms, is operated by the University of Illinois to help launch successful startup companies.

HPCBio

The High-Performance Biological Computing (HPCBio) research facility provides a single point of access, open to researchers from all campus units, helping them to find solutions to their biomedical data management and analysis problems. It is a collaborative effort between multiple partners in the University community. It is administratively attached to the Roy J. Carver Biotechnology Center, and incorporates the Center's Bioinformatics Unit. It is strongly anchored in the genomics research agenda of the Institute for Genomic Biology. It builds on the research expertise and facilities IGB, its faculty and affiliates. The National Center for Supercomputing Applications (NCSA) provides high-level technical expertise in high-performance computing, and the Office of the Vice-Chancellor for Research provides financial support and institutional oversight. Director Dr. C. Victor Jongeneel is with NCSA; eight other staff members are listed.

Its missions are:

- To provide an infrastructure for bioinformatics, combining hardware, software, databases, training, consulting and services;
- To perform applied research in biological computing with a view to developing and maintaining innovative, robust, and validated software that leverages both the latest data generation technologies and advances in computer architectures.

UNITS WITH JOINT EDUCATIONAL AND RESEARCH FUNCTION

Computational Science and Engineering (CSE)

The Computational Science and Engineering program at University of Illinois was established by the College of Engineering to foster interdisciplinary research and education. The program boasts 245 faculty affiliates across 27 departments. Research is supported by providing students and faculty access to computing resources and also, through a strong training program. This training program offers workshops and short courses, generally with open enrollment for maximum impact, on computational tools and techniques used in modern engineering and science. Current and upcoming workshops are listed on http://cse.illinois.edu/training-technical-computing-workshops. Topics include MATLAB, Mathematica, ANSYS FEA, Fluent, ABAQUS, Numerical Python, and Big Data Analytics.

In its educational mission, CSE supports a popular certificate program for students at both the graduate and undergraduate level. This program establishes a set of coursework, spanning disciplines that students must complete in addition to typical requirements in their major. Current enrollment is around 200 students, mostly at the PhD level and comprised of some of the best students at one of the top Engineering universities in the world.

CSE makes parallel computing resources available to students, faculty, and courses through the Illinois CampusCluster. Located in the Advanced Computational Building, the CSE Parallel Computing Resource consists of approximately 300 CampusCluster compute nodes, each with two 2.67 GHz Intel Xeon hex-core processors and 24 GB of RAM, for a total of about 3600 cores. The primary network connecting the cluster nodes is Voltaire QDR Infiniband. In addition, all nodes in the cluster are also connected by a 1 Gbs switched, full-duplex Ethernet. The operating system for the cluster is Scientific Linux. The cluster can be accessed by approved users only via secure shell (ssh or slogin) to the front-end array, taub.campuscluster.illinois.edu, with your University NetID and AD password. More details on getting access can be found at http://cse.illinois.edu/research/taub/

Technology Entrepreneur Center

An interdisciplinary center that fosters entrepreneurship through its courses, venture and product competitions, workshops, and events. TEC offers on-site and online certificate programs for education and professional development and hosts outreach activities for students and alumni. About a thousand students take a course in the TEC center annually.

Illinois Informatics Institute

An interdisciplinary institute devoted to informatics-related research and education programs across campus, encouraging research projects that involve applications of computing and information in science, arts, and humanities. The Institute has a PhD program and a bioinformatics MS degree. The center includes 120 faculty affiliates in many departments, 30 PhD students, and seven staff members working from NCSA and Lincoln Hall.

Center for Biophysics and Computational Biology

An interdisciplinary center of over 40 faculty members that serves physical and computer science students who are interested in applying their knowledge to biology as well as students with a biological background interested in instrumentation, computation, and physical aspects of biology. Current research uses a wide range of computer platforms to simulate diverse biological phenomena at many levels. Center members are also developing bioinformatics tools to create and search biological databases to provide input for functional analysis and simulations.

SIZEABLE INTERDISCIPLINARY GROUPINGS OF FACULTY RESEARCHERS

Center for Informatics Research in Science and Scholarship

Associated with the Graduate School of Library and Information Sciences, the center conducts research on information problems that impact scientific and scholarly inquiry. Projects and activities focus on how digital information can advance the work of scientists and scholars, the curation of research data, and the integration of information within and across disciplines and research communities.

Core research areas are:

- Socio-Technical Data Analytics
- Digital Collections and Curation
- e-Science;
- Digital Humanities

Information Trust Institute

The ITI is an academic and industry partnership that conducts foundational and applied research in trustworthy and secure information systems. It recently celebrated its ten year anniversary.

Parallel Computing Institute (PCI)

The Parallel Computing Institute aims to provide researchers and educators the support needed to address major computational challenges in science, engineering, health, and business, and to build powerful and efficient infrastructure for computation-heavy applications. Forty-five research faculty are associated with the Institute, which is led by Director Bill Gropp and Chief Scientist Wen-mei Hwu. PCI initiatives include:

- Inpatient MRI,
- DNA sequencing,
- Productive heterogeneous exascale computing
- Data layout transformations,
- Scalable parallel programming.

KnowEnG: An NIH Center of Excellence in Big Data Computing

KnowEnG is a National Institutes of Health-funded initiative that brings together researchers from the University of Illinois and the Mayo Clinic to create a Center of Excellence in Big Data Computing. It is part of the Big Data to Knowledge (BD2K) Initiative that NIH launched in 2012 to tap the wealth of information contained in biomedical Big Data. KnowEnG is one of 11 Centers of Excellence in Big Data Computing funded by NIH in 2014.

This four-year project will create a platform where biomedical scientists, clinical researchers, and bioinformaticians can bring their own data and perform common as well as advanced analysis tasks, guided by the "knowledge network", a large compendium of public-domain data. The knowledge network embodies community data on genes, proteins, functions, species, and phenotypes, and relationships among them. Instead of analyzing their data set in an isolated fashion, researchers will be able to go straight to asking global questions. The infrastructure, capacity and tools will grow with the datasets.

NVIDIA Cuda Center of Excellence

In 2008, the University of Illinois at Urbana-Champaign was named the world's first CUDA Center of Excellence by the NVIDIA Corporation, a world leader in visual computing technologies. As an NVIDIA Center of Excellence, Illinois is recognized as a primary collaborator with NVIDIA, and has received research funding and equipment to expand that working relationship. Recent NVIDIA support has included providing 32 QuadroPlex Model 4 systems, each outfitted with 64 graphics processing units (GPUs). These systems have been incorporated into a 16-node CUDA cluster, which is managed by NCSA as a key component of Illinois's parallel computing facilities. The Center is affiliated with the Parallel Computing Institute. The Principal Investigator of the first CUDA Center of Excellence is Prof. Wen-mei Hwu.

Health Information Technology Center

The Health Information Technology Center (HITC) at the University of Illinois Urbana and Chicago campuses promotes research on using information technologies to support health and healthcare. Areas of interest for HITC include: data sciences, privacy and security, usability, and safety. These areas are overlapping in scope and include issues related to both primary and secondary use of health information. The Center has 35 affiliated faculty members at Urbana and Chicago, fellowships, and seminars. Director is Prof. Carl Gunter.

CyberGIS Center

The CyberGIS Center for Advanced Digital and Spatial Studies (CyberGIS Center) was established in 2013 as a partnership among several units at the University of Illinois at Urbana-Champaign. CyberGIS represents a vibrant interdisciplinary field for bridging advanced cyberinfrastructure, geographic information science and technologies, and various geospatial knowledge domains. The mission of the CyberGIS Center is to empower advanced digital and spatial studies through innovation of CyberGIS technologies and applications. The Center has about twenty faculty affiliates from Computer Science, Civil and Environmental Engineering, the College of ACES, GSLIS, NCSA, and other departments. The Center has a biweekly brownbag lunch and sponsors a distinguished lecture series. Dr. Shaowen Wang is the Founding Director of the CyberGIS Center.

Initiative for Mathematical Sciences and Engineering (IMSE)

The long term goal of IMSE is to become a national Center for the interaction of mathematics and engineering. IMSE will accelerate advances and innovations in mathematics and engineering through cross-fertilization, with the dual objectives of fostering cutting-edge mathematics as indispensable in addressing signi cant engineering applications and of advancing the next generation of mathematics through the infusion of new classes of problems. IMSE will serve as a catalyst

for mathematicians and engineers to work together on such problems, by engaging the former in emergent engineering applications and equipping the latter with state-of-the-art mathematical tools.

Center for Exascale Simulation of Plasma Coupled Combustion (XPACC)

The Center for Exascale Simulation of Plasma-Coupled Combustion (XPACC) will be working to develop a new mode of managing combustion and aims to make breakthroughs in this emerging field at the basic science level. It is funded for five years by a $16 million grant from the National Nuclear Security Administration, part of the U.S. Department of Energy. The center, one of three Multidisciplinary Simulation Centers funded through NNSA's Predictive Science Academic Alliance Program II, comprises of researchers from Illinois and the Ohio State University. The Illinois researchers hail from aerospace engineering, chemistry, computational science and engineering, computer science, electrical and computer engineering and mechanical science and engineering. The Ohio State investigators are experts in plasma-assisted combustion, kinetics and experimental diagnostic techniques.

RESEARCH GROUPS OF ONE TO TEN FACULTY MEMBERS

Materials Computation Center

The Materials Computation Center (MCC), a collaboration among five faculty members, is engaged in developing software to analyze and predict the properties of materials, training students in computational materials science, and involving undergraduates in experimental and computational research. It supports educational experiences in materials science through the Travel Award Program in conjunction with European organizations Psi-k and CECAM. It also supports annual summer schools on computational materials science and conferences.

Theoretical and Computational Biophysics Group

An interdisciplinary research group with seven faculty members, directed by Professor Klaus Schulman, is supported by the National Institutes of Health, the National Science Foundation, and other agencies that models large macromolecular systems in realistic environments to understand biomolecular processes coupled with mechanical force, bioelectronic processes in metabolism and vision, and the function and mechanism of membrane proteins. Ongoing initiatives include:

- Molecular modeling tools that integrate structural information with bioinformatics databases and molecular dynamics simulations;
- High performance molecular visualization and simulation software modelling biomolecules in realistic environments of 100,000,000 atoms or more;
- Conceptual and methodological foundations of molecular modeling in the fields of quantum biology, mechanobiology, and interactive modeling;
- Biomedical science through collaborations between theoretical and experimental researchers;
- Support of search process and training through a web-enabled collaborative environment;
- Service, training, and dissemination by leveraging web-based molecular graphics and integrated modeling technologies.

Text Information Management and Analysis Group (TIMAN)

A research group in the Department of Computer Science headed by Prof. Cheng Xiang Zhai, is working on developing general technologies for managing and exploiting large amounts of text data and applying such technologies to multiple application domains such as health and medical informatics, bioinformatics, education, and Web data mining. Research directions include

- Intelligent search engine algorithms
- Personalization of search and information service
- Conversational search systems
- Robust and general natural language processing
- Text data analytics
- Biomedical and health informatics
- Intelligent education systems

Cognitive Computation Group

The research group, headed by Prof. Daniel Roth, focuses on the computational foundations of intelligent behavior, attempting to understand the role of learning in supporting intelligent inference and use this understanding to develop systems that learn and make intelligent inferences in complex domains. The group's work spans foundational questions in learning, knowledge representation and reasoning to experimental paradigms and large scale system development and draws on methods from theoretical computer science, probability and statistics, artificial intelligence, linguistics and experimental computer science.

OTHER RESEARCH CENTERS

Institute for Computing in Humanities, Arts, and Social Science (I-CHASS)

This institute fosters collaboration between humanists, artists, and social scientists and researchers in computer science, engineering, and high performance computing. It emphasizes identifying, creating, and adapting computational tools that accelerate research and education.

HathiTrust Digital Library and Research Center

HathiTrust is a partnership of major research institutions and libraries working to ensure that the cultural record is preserved and accessible long into the future. There are more than 80 partners in HathiTrust, and membership is open to institutions worldwide.

The HTRC is a collaborative research center launched jointly by Indiana University and the University of Illinois, along with the HathiTrust Digital Library, to help meet the technical challenges of dealing with massive amounts of digital text that researchers face by developing cutting-edge software tools and cyberinfrastructure to enable advanced computational access to the growing digital record of human knowledge.

The Cline Center for Democracy

The Cline Center for Democracy at the University of Illinois Urbana-Champaign aims to promote human well-being by creating new knowledge to better understand and address key social issues in the contemporary world. The Cline Center does this by empowering a new generation of interdisciplinary social science research that uses data science methods to study a wide range of contemporary social issues on a worldwide scale, to publicize the results of that research, and to stimulate broader interest in solving those issues.

One of the Cline Center's signature projects is the Social, Political and Economic Event Database (SPEED) project. SPEED is a technology-intensive effort to extract data from news reports about small-scale civil unrest events like protests and acts of political violence, as well as governmental responses to those activities. SPEED documents civil unrest activity for every country in the world from World War II to the present using a global archive of news reports. Within SPEED, event data are generated by a hybrid system combining fully-automated machine-learning and natural language processing technologies with human analysts who draw from a suite of sophisticated tools to implement carefully structured and pretested protocols.

eDream - the Illinois Emerging Digital Research and Education in Arts Media Institute

eDream is an institute dedicated to promoting arts that are conceived, created, and conveyed through digital technologies. It brings together artists, scholars, and technical innovators from many disciplines, diverse perspectives to realize the dream of arts futures in present practice.

Center for IT and e-Business Management (CITEBM)

An interdisciplinary center associated with the College of Business, developing world-class research and educational programs on electronic commerce, information technology strategy, technology management, and the commercialization of technology. It brings together facilities and researchers in the fields of business administration, economics, accounting, finance, computer science, psychology, and engineering, CITEBM is aimed at developing new managerial insights that will lead to effective business systems, strategies, processes, and models for commerce and business administration. Research areas include:

- Electronic commerce, information technology strategy, technology management, and the commercialization of technology,
- Global electronic commerce,
- Business models and strategies for applying new technology,
- Multidisciplinary approaches in the research and educational programs that reflect managerial, technological, and sociological perspectives.

Center initiatives include establishing a business incubator and building strong ties with the industrial and business communities to collaborate on developing state-of-the-art research and education programs.

RECENTLY COMPLETED PROJECTS

Multimodal Information Access and Synthesis Center

The Center, supported by the Department of Homeland Security, studied the questions of how to deal with the huge amount of unstructured data as if it was organized in a database with a known schema--how to locate, organize, access and analyze unstructured data. Its goals are to develop fundamental theories, computational models, algorithms, and tools to enable analysts to transform raw data into useful and understandable information.

Illinois Cloud Computing Testbed

The Illinois Cloud Computing Testbed (CCT) is the world's *first cloud testbed* aimed at supporting *both* systems innovation *and* applications research within a single microcosm. It was founded by CS Professors Campbell, Gupta, and Heath.   CCT is unique in several respects: (1) it is a true cloud/datacenter testbed, e.g., its storage to computation ratio is different from that of many existing testbeds such as Emulab and PlanetLab). Currently, CCT is configured with about 500 TB of shared storage and 1000+ shared cores; (2) it is the only cloud testbed to support both applications and systems research (in contrast with the Google-IBM testbed, for example).

Research efforts on CCT, both internal to the University of Illinois and external, included the following research areas: networking, operating systems, databases, storage, virtual machines, distributed systems, data-mining, Web search, network measurements, and multimedia.

Universal Parallel Computing Research Center

An interdisciplinary research center that was designed to enable Illinois researchers from across campus to come together in new, application-focused research centers and achieve their scientific goals using the latest and most efficient parallel computing technologies. With funding and support from Microsoft and Intel, UPCRC Illinois paved the way for mass-market systems to utilize supercomputing parallel programming techniques.

The Center for Process Simulation and Design

The Center for Process Simulation and Design (CPSD) was dedicated to improving the quality and performance of products and materials through simulation and optimization of the industrial processes by which they are manufactured. Its initial focus was on understanding and controlling microstructural material properties resulting from casting and extrusion processes. Such problems have multiple length and time scales, moving boundaries, and complex, dynamically evolving geometries and topologies. This interdisciplinary program included a broad range of research activities in engineering, mathematics, and computer science. Ten faculty from Computer Science, Physics, and Mathematics were affiliated with the Center, which completed its work in 2007.

Appendix B: Computing Events and Symposia on Campus

Big Data Summit: http://researchpark.illinois.edu/bigdatasummit
November 5, 2014

IMSE Hot TIME Symposium: https://imse.math.illinois.edu/?page_id=603
Feb. 24 & 25, 2013

John Deere Big Data Summit
Oct. 1 & 2, 2013

International Digital Curation Conference: http://www.dcc.ac.uk/events/idcc14
Feb. 24-27, 2014

Illinois Research Data Initiative Events:
http://blogs.cites.illinois.edu/datasteward/events-2/
Fall 2011 & Spring 2012

Scholarly Commons Open Access Week:
http://publish.illinois.edu/commonsknowledge/2013/09/03/open-access-week-october-21-25/
Oct 21-15, 2013

Annual Allerton Conference on Communication, Control, and Computing:
http://publish.illinois.edu/cslallertonconference/
Late September - early October each year

CSL Symposium on Emerging Topics in Control and Modeling:
http://publish.illinois.edu/emerging-topics-2013/
Nov. 7 & 8, 2013

Blue Waters Symposium:
https://bluewaters.ncsa.illinois.edu/symposium-may-2014
May 12-15, 2014

Appendix C: Peer Institutes and Consortiums

Great Lakes Consortium for Petascale Computation:

http://www.greatlakesconsortium.org/about.html
The Great Lakes Consortium for Petascale Computation is a collaboration among colleges, universities, national research laboratories, and other educational institutions. The consortium facilitates the widespread and effective use of petascale computing, through the development of new computing software, applications, and technologies. A comprehensive educational and workforce development program ensures that advances made by consortium members are passed on to the next generation of researchers and applied to frontier research questions in science, technology, engineering, and the social sciences.

UC Berkeley:

http://datascience.berkeley.edu/about/about-the-uc-berkeley-school-of-information/

UC Berkeley researchers at the forefront of data science, as evidenced by the recent creation of the Social Sciences Data Laboratory (D-Lab) for data-intensive social science research; the AMPLab (Algorithms Machines People), which focuses on machine learning; the Simons Institute for the Theory of Computing; and a Masters of Data Science program in the School of Information.

A new center, The Berkeley Institute for Data Science, was announced in November. It will be housed in the campus's central library building, is made possible by grants from the Gordon and Betty Moore Foundation and the Sloan Foundation, which together pledged $37.8 million over five years to three universities – UC Berkeley, the University of Washington and New York University – to foster collaboration in the area of "data science."

http://citris-uc.org/about-citris/

The Center for Information Technology Research in the Interest of Society (CITRIS) was formed in 2001, when researchers within the UC system realized that the real opportunities lay not just in developing new and innovative technologies, but in *applying* them. CITRIS was established to address the most pressing social and environmental issues facing California. To meet this goal, they focus our research on four core initiatives: Energy, Health Care, Intelligent Infrastructure, and Data and Democracy.
- The i4Energy Initiative employs sensors and emerging technology to gather, manage, and utilize information about energy conservation and expenditure from the household-level to the grid.
- The Health Care Initiative aims to improve access and reduce disparities and costs in health care across the state by developing and integrating technology advances in telehealth, sensors, services and gaming.
- The Intelligent Infrastructure Initiative addresses the areas of water, cities, and transportation to develop and deploy intelligent "cyber-physical" systems to better manage scarce resources and promote sustainability.
- The Data and Democracy Initiative creates tools to foster public engagement in critical social, political, and economic issues by exploring the dynamic relationships between digital media and democratic practices.

CITRIS INVENTION ECOSYSTEM – *Learn, Build, Launch, Connect*
CITRIS brings a unique methodology and collection of resources to its pursuit of innovation across disciplines and subject areas. With our cutting-edge laboratories and renowned faculty experts, CITRIS serves as an in-

cubator for translating new ideas into working prototypes. This iterative process has four steps: *Learn, Build, Launch, Connect.*
As a university-based research center, it is committed to furthering UC's educational mission and creating opportunities to *Learn* emerging skills.

With the Berkeley Marvell NanoLab, The Foundry at CITRIS startup incubator, and the CITRIS Invention Lab, CITRIS offers unparalleled facilities to *Build* prototypes and demonstration models for projects from the nanoscale to urban scale. Once a proof of concept has been tested and refined, it is ready for *Launch* to industrial partners, public agencies, and other potential users. The final step is to *Connect* these innovations with the networks that will benefit from and promote their implementation. As the effectiveness of each device or intervention is evaluated, the process begins again as we *Learn* for the next iteration.

CITRIS has pioneered this model in the university setting, providing an "invention infrastructure" that creates physical and online products along with the processes to support continuous innovation. The combination—information technology research with state-of-the-art facilities creating products for the benefit of society—distinguishes CITRIS from other interdisciplinary research centers.

UC San Diego:

Research Cyberinfrastructure (RCI):
http://rci.ucsd.edu/
RCI is a UCSD-sponsored program that offers campus researchers the facilities, storage, data curation, computing, networking, and expertise to facilitate their research using shared cyberinfrastructure services across campus.
The RCI program provides UCSD principal investigators these cost-effective, reliable services for use in ongoing research and for incorporation into future research proposals. In general, these services are available at a reduced cost, subsidized by the RCI program. All services are also available to non-UCSD researchers at attractive non-subsidized rates.
RCI has a complete set of research infrastructure services, described in detail and available through this website. Most of these services are in production, while some continue in pilot phases to best determine researcher requirements and appropriate business models.

Carnegie Mellon University:

Computing, IT and Security:
http://www.cmu.edu/research/projects/computing.shtml
At Carnegie Mellon, computing has always been more than software, code, usernames and passwords. It's about robots exploring the planetary terrain, making a game that teaches middle-schoolers how to program and establishing a private-public partnership to create the most secure systems in the world. Here's a snapshot of some of the interesting research happening at our university.

Quality of Life Technology Center:
http://www.cmu.edu/qolt/

The Quality of Life Technology Center is a National Science Foundation (NSF) Engineering Research Center (ERC) focused on the development of intelligent systems that enable older adults and people with disabilities to live more independently.

University of Michigan:

http://arc.research.umich.edu/

Advanced Research Computing at U-M provides access to and support for the use of advanced computing resources. Rapid developments in technology are creating unprecedented capabilities for handling massive amounts of data. ARC helps open up these new and more powerful approaches to research challenges in fields ranging from physics to linguistics, and from engineering to medicine.
ARC brings researchers and advanced computing resources together through a number of services and resources:

- **A shared computing cluster, Flux,** operated on an allocation basis. Researchers can buy as much or as little time as needed on our cluster, currently more than 12,000 cores and growing.
- **Events for individuals engaged in computational discovery** at U-M. These include an annual conference, CI Days, an occasional speaker series, and education and training sessions.
- **Access** to regional and national high-performance computing resources, as well as on-campus infrastructure.

Michigan State University:
Institute for Cyber-Enabled Research:
http://icer.msu.edu/
The Institute for Cyber-Enabled Research (iCER) provides a common structure for researchers from across academia and industry to work on how computation can better their research
The High Performance Computer Center (HPCC) at Michigan State provides computational hardware and support to MSU faculty, students and researchers:

- Graphical Processing Units (GPUs)
- Large Shared Memory Systems (Fat Nodes)
- High Throughput HTCondor Cluster
- General Purpose Clusters

University of Minnesota:
https://www.msi.umn.edu/

The Minnesota Supercomputing Institute seeks to provide researchers at the University of Minnesota access to high-performance computing resources and user support to facilitate successful and cutting-edge research in all disciplines, help researchers attract funding, contribute to undergraduate and graduate education, and benefit the broader community.
MSI is committed to expanding and developing the types of service it offers in order to continue to play its key support role across the growing spectrum of scientific fields.
MSI is also committed to facilitating University-industry collaboration and to promoting technology transfer through the interchange of ideas in the field of supercomputing research, including the dissemination of results of research accomplished with MSI resources.

Georgia Tech:

Institute for Data and High Performance Computing:
http://www.hpc.gatech.edu/
Big Data Research:
http://www.research.gatech.edu/areas/big-data
Through a number of campus units, including the Institute for Data and High Performance Computing(IDH) and the Georgia Tech Research Institute (GTRI), Georgia Tech supports multidisciplinary research teams that are both developing innovations in computational methods to advance Big Data analysis, and applying these techniques to industry, business, and the public sector. Enabling technologies under development include data visualization, advanced analytics, machine learning, and high-performance computing. Application areas for Georgia Tech's Big Data research include astrophysics, biomedicine, combustion, energy, finance, health care, manufacturing, materials, information and cybersecurity, social networks, sustainability, and transportation. Both undergraduate and graduate students contribute to research in these critical areas.

CalTech:

http://www.cacr.caltech.edu/main/
The mission of the Center for Advanced Computing Research (CACR) is to ensure that Caltech is at the forefront of computational science and engineering (CSE). CSE is the practice of computer-based modeling, simulation, and data analysis for the study of scientific phenomena and engineering designs. Computer modeling and simulation make it possible to investigate regimes which are beyond current experimental capabilities and to study phenomena that cannot be replicated in laboratories, such as the evolution of the universe. The results often suggest new experiments and theories. Computation is also essential for processing the flood of high-dimensional data generated by modern instruments.
CACR provides an environment that cultivates multidisciplinary collaborations. CACR researchers take an applications-driven approach and currently work with Caltech research groups in aeronautics, applied mathematics, astronomy, biology, engineering, geophysics, materials science, and physics. Center staff have expertise in data-intensive scientific discovery, physics-based simulation, scientific software engineering, visualization techniques, novel computer architectures, and the design and operation of large-scale computing facilities.

UT Austin:

https://www.ices.utexas.edu/
The Institute for Computational Engineering and Sciences (ICES) is an organized research unit created to foster the development of interdisciplinary programs in computational sciences and engineering (CSE), mathematical modeling, applied mathematics, software engineering, and computational visualization. The Institute currently supports eleven research centers and alliances and six research groups, and with additional units still in the planning stages. It also supports seven major programs including the following:

- The **Computational Science, Engineering, and Mathematics Program (CSEM)** is a graduate degree program leading to the MS and PhD degrees in CSEM.
- The **ICES Postdoctoral Fellowship Program** supports outstanding computational scientists who have recently completed doctoral studies in an area relevant to research conducted at the Institute.
- The **J. Tinsley Oden Faculty Fellowship Research Program** brings outstanding researchers and scholars from academia, government laboratories, and industry to collaborate with ICES faculty and students.
- The **Moncrief Endowed Positions Program** is dedicated to recruiting outstanding junior faculty committed to research and academic work in CSE.
- The **Moncrief Grand Challenge Awards Program** provides resources for UT Austin faculty to work on the Grand Challenges in CSE that effect the competitiveness of the nation.
- The **Moncrief Undergraduate Summer Internship Program** supports qualified undergraduate students to work with ICES faculty on ongoing research initiatives during the summer months.

- ICES **Undergraduate Certificate Program** in Computational Science and Engineering provides junior and senior level students at UT Austin an opportunity for in-depth study in CSE.

MIT:

http://www.csail.mit.edu/

The Computer Science and Artificial Intelligence Laboratory – known as CSAIL – is the largest research laboratory at MIT and one of the world's most important centers of information technology research.

CSAIL and its members have played a key role in the computer revolution. The Lab's researchers have been key movers in developments like time-sharing, massively parallel computers, public key encryption, the mass commercialization of robots, and much of the technology underlying the ARPANet, Internet and the World Wide Web.

CSAIL research is focused on developing the architectures and infrastructures of tomorrow's information technology, and on creating innovations that will yield long-term improvements in how people live and work. Lab members conduct research in almost all aspects of computer science, including artificial intelligence, the theory of computation, systems, machine learning, computer graphics, as well as exploring revolutionary new computational methods for advancing healthcare, manufacturing, energy and human productivity.

Stanford:

https://doresearch.stanford.edu/research-scholarship/computing-support-research#the-stanford-research-computing-facility

Research Computing at Stanford is a joint effort of the Dean of Research and IT Services to build and support a comprehensive program to advance computational research at Stanford. That includes traditional high performance computing (HPC) as well as high throughput and data – intensive computing.

The Stanford Research Computing Facility (SRCF) provides the campus research community with data center facilities designed specifically to host high performance computing equipment. Supplementing the renovated area of the Forsythe data center, the SRCF is intended to meet Stanford's research computing needs for the coming years. A Stanford building located on the SLAC campus, the SRCF was completed in the fall of 2013, with production HPC services being offered as of November 2013. The facility and services therein are managed by the Stanford Research Computing Center (SRCC).

Rensselaer Polytechnic Institute:

http://idea.rpi.edu/

RPI has invested 60 million in The Rensselaer IDEA: Harnessing the Power of Data to Change the World, led by Jim Hendler: http://news.rpi.edu/content/2013/06/13/rensselaer-idea-harnessing-power-data-change-world#sthash.kqRz4b8S.dpuf. The Rensselaer Institute for Data Exploration and Applications (IDEA) will enable research across their campus to access technologies via the development of critical computational methodologies including data-intensive supercomputing, large-scale agent-based simulation, and cognitive computing technologies.

University of Washington:

http://escience.washington.edu/

In November, the University of Washington announced a new 5-year partnership with UC Berkeley, NYU, the Gordon and Betty Moore Foundation, and the Alfred P. Sloan Foundation to build a distributed Data Science Environment to advance data science research and practice across all disciplines. The data science initiative is part of the eScience Institute led by Ed Lasowska.