

IMPROVING SOCIAL SCIENCE RESEARCH COMPUTING AT NORTHWESTERN UNIVERSITY

July/October 2002

Committee Members

Joe Altonji (WCAS Economics, Institute for Policy Research)
Frank Cervone (NU Library, IT)
Greg Duncan (SESP, Institute for Policy Research)
Cathy Grimsted (KSM Administration)
Shelby Haberman (WCAS Statistics)
Madelyn Iris (FSM Medicine, Buehler Center on Aging)
Ann Janda (NU Library Social Science Data Services)
Bob Korajczyk (KSM Finance)
Bobai Li (WCAS Sociology)
Anthony T. Lo Sasso (Institute for Health Services Research and Policy Studies)
Ruth Reingold (WCAS Computing Technology)
Cynthia Robin (WCAS Anthropology)
Wes Skogan (WCAS Political Science, Institute for Policy Research)
Ken Woo (Law Computing)

Fred Arnold (NUIT AT Research Technologies)
Bruce Foster (NUIT AT Research Technologies)
Gary Greenberg (NUIT Teaching & Research Initiatives)
Tom Halloran (NUIT Computer Services)
Bob Taylor (NUIT Academic Technologies)

IMPROVING SOCIAL SCIENCE RESEARCH COMPUTING AT NORTHWESTERN UNIVERSITY

SUMMARY	1
INTRODUCTION	2
The DCE/DFS Cell	2
Social Science Research Computing at Northwestern University	3
FUNCTIONAL AND NEEDS REQUIREMENTS	3
Computation	3
Applications	4
Network Storage	4
Education and Consulting	5
RECOMMENDATIONS	6
Migrate the Current DCE/DFS Cell to Intel-based Systems Running Linux	6
Create a Linux Cluster for Computationally Intensive Research	7
Support Windows-based Social Science Research Computing	7
New Applications: Geographic Information Systems (GIS)	8
CONCLUSIONS	9
APPENDIX A: SOCIAL SCIENCE RESEARCH COMPUTING AT OTHER UNIVERSITIES	11
The University of Chicago	12
Stanford University	13
University of Michigan – Ann Arbor	14
The University of Wisconsin – Madison	15

SUMMARY

After preliminary meetings in the Fall, 2001 with department representatives to confirm the need for reviewing social science research computing and identifying an initial set of issues and concerns, a committee was formed in January, 2002 to review needs, define functional requirements, and consider possible solutions.

The process began with the assumption that the current environment, the DCE/DFS Cell, had reached the limits of its capacity and, with diminishing vendor support, would have to be replaced. The Committee discovered that commercial demand and support for DCE has revived significantly and cluster computing, which is receiving much attention, is not easy enough to use for general research computing. The Committee also found that standard enterprise approaches to computation and storage may be too costly to be sustained and have potential security and performance problems for social science research. The Committee was also interested in an approach that could provide a sustainable funding model for the ongoing maintenance and expansion of central as well as local resources.

The Committee makes four recommendations:

- Migrate the current DCE/DFS Cell to lower-cost Intel-based systems running Linux to expand computation and storage capacity and address unmet demand with minimal disruption to current research and instruction.
- Create a small cluster of Intel-based computers running Linux for technically sophisticated researchers with computationally intensive problems not easily addressed by the Cell.
- Support Windows-based social science research computing by providing network storage for datasets and research data.
- Expand the suite of available applications in order to provide new areas of opportunity for social science research computing, especially Geographic Information Systems (GIS).

The Committee believes that following these recommendations will result in a robust social science research computing environment that will attract and retain faculty and students, support larger, more diverse research communities, remove many of the existing constraints on what types of problems can be addressed by faculty and their graduate students, and make it possible to enhance undergraduate social science education to better prepare students for careers and graduate research.

INTRODUCTION

The DCE/DFS Cell

In 1993 Vice President Rahimi charged a faculty advisory committee to identify the needs of social science research computing at Northwestern University (<http://www.it.northwestern.edu/committees/sscac.html>). In response to the report, Information Technology funded replacing a mainframe computer that had been supporting social science research computing. In 1995 a new Distributed Computing Environment (DCE) with a Distributed File System (DFS), the DCE/DFS Cell, went into service. The Cell was designed to make more high-performance resources available than faculty, departments, and programs could provide on their own and establish a common, well-managed network environment that integrated central and local resources.

Currently, the DCE/DFS Cell (<http://www.at.northwestern.edu/rtg/soc-sci.ssi>) consists of seventeen computation servers, two DFS file servers, a DFS fileset location database server, two DCE security/directory servers, and a magnetic tape server. The DFS file service provides approximately 144 GB of shared storage. Files are usually compressed to 5% of their original size, providing the equivalent of 2TB storage. All but two of the compute servers and much of the storage in the Cell were purchased by specific departments or programs, often for restricted use. Site-licensed statistical applications are readily available from the Cell. The DCE/DFS Cell supports 725 user accounts that have access to the two primary compute servers. No matter which shared or restricted compute server is being used, the users' home directories are the same. Even as capacity and the number of users has increased since the Cell was created, one person continues to manage storage, backup services, performance, and applications over the multiple servers in the Cell. This network-based approach to supporting research is admired by social scientists outside the University.

Today, the network storage is saturated and the servers are aging and need to be replaced. The software used to manage the environment has not been well-supported commercially in recent years. Furthermore, the low-cost Intel platform has caught up with and even surpassed the performance of traditional UNIX platforms for single-user situations. In addition to replacing central resources, faculty, departments, and programs are looking for direction on how they should replace their own equipment in the Cell to address their growing needs for computation and storage.

With the University's social science research computing environment aging and having reached its capacity, there is growing concern that Northwestern may not be able to attract potential faculty and graduate students because it lacks a robust computational environment for social science research. Equally important, current faculty may respond to recruiting efforts by other universities that can offer better computing resources and more extensive support. There is also interest in being able to offer opportunities for undergraduates to learn computational methods that will better prepare them for careers and graduate programs. What is done now to improve the state of social science research computing will be important for faculty recruitment and retention, graduate student recruitment, and the quality of social science research and instruction at the University.

Social Science Research Computing at Northwestern University

Social science research computing makes important contributions to the University's research and education mission. This is particularly true for faculty research, graduate student work, and undergraduate courses in the Weinberg College of Arts and Sciences, the Kellogg School of Management, the School of Education and Social Policy, and the Feinberg School of Medicine, as well as research at the Institute for Policy Research, the Joint Center for Poverty Research, the Institute for Health Services Research and Policy Studies.

Research computing resources are also important for students. Undergraduates learn how to analyze data through class assignments, the Mathematical Methods in the Social Sciences program, or NU Library Data Services. Graduate students may work with faculty on supervised projects before moving on to their own dissertation research. While the scope of their problems may vary, undergraduate and graduate students share the same needs as faculty for applications, computation, storage, and support.

Overviews of social science research computing at The University of Chicago, Stanford University, the University of Michigan – Ann Arbor, and The University of Wisconsin – Madison are included at the end of this report in Appendix A.

FUNCTIONAL AND NEEDS REQUIREMENTS

Social science research computing involves reading and processing data files, some as large as 1GB, running simulations, developing custom programs, and presenting the results of analysis. Extracted datasets are often saved for use as long-term work files. Statistical and mathematical software is used to analyze the data. The larger the model or the more complex the simulation, the longer computation will take. The faster the processor or the more processors available, the faster data can be analyzed and the more complex the model or simulation can be. Large amounts of network storage may be needed to extract and analyze source data. For better performance, local scratch space avoids moving data back and forth across the network. The primary requirements for social science research computing are computation, applications, network storage, and education and consulting.

Computation

Computational resources determine the size and scope of the problems one might address. As datasets grow and questions become more complex, computation must keep up. Simulation and modeling nonlinear problems can require days, even weeks of computation and large amounts of memory.

- Central computation resources must meet the demands for processing very large datasets, as well as running complex simulations and custom programs. Inadequate computation can make working with large datasets unwieldy and simulations intractable.

- Central computational resources should be scalable so they can meet growing needs and expectations. Individuals and research groups that might otherwise make excessive demands on shared resources should be able to add their own computational resources.
- Shared computational resources must keep up with advances in performance in order to address the increasingly intensive computational needs of faculty and their students.
- The computational (and storage) needs of graduate students must be considered to assure they have adequate resources for experimentation without having to scale back the scope of their investigations.

Applications

Ready access to statistical applications to analyze data, simulation tools to explore increasingly complex models, and programming resources to develop specialized tools is essential. A suite of applications is required to meet the varied needs of social scientists.

- Necessary application software includes SAS, Stata, MATLAB, GAUSS, and Stat/Transfer (see http://charlotte.at.northwestern.edu/dce/bull/stat_sw.html)
- Software such as Geographic Information System (GIS) and qualitative analysis software that offer new opportunities for social science research and provide useful tools for fields that have not previously relied on computational methods should also be made available.
- Whenever demand and cost permit, software should be site-licensed. If possible, licenses should extend to student computers, as well as faculty and departmental computers.
- Software should be maintained and shared from a central server over the network. Rather than having to install and maintain software themselves, users should be able to run applications over the network, knowing that the most recent version is always available.
- Applications should be readily accessible across the campus network from public and department computer labs where students are likely to work on course assignments and independent research projects.

Network Storage

Datasets are at the core of social science research computing and network storage is as important as computation. Not only must there be enough storage to hold very large datasets, there must be additional temporary storage for extracting and analyzing data. Once data has been extracted, they are usually saved to avoid having to extract the data each time they are needed.

- Data files should be stored on the network so they can be used from different locations and shared with collaborators both on- and off-campus.
- While quotas on permanent network storage can control normal use, there should always be enough temporary “scratch” storage for people to get their work done and experiment.
- Storage should be easily expandable. Individuals and groups should be able to add their own storage, especially if meeting their needs would interfere with other users.
- Security requirements for using confidential data should be addressed as widely as possible so faculty do not have to work on isolated systems.
- Storage, backup, and recovery should be managed transparently for users over the network. Users should not have to worry about where their data is stored.

Education and Consulting

The consequence of inadequate training and support is decreased productivity, increased frustration, poor use of resources, and the inability to effectively take advantage of consulting. Consulting and education services are needed by faculty, the graduate students working under them, and the undergraduates taking their courses. Opportunities are lost, both for good research and for funding.

- Consulting on statistical methods, using applications, programming, and system administration is very important for both faculty and students.
- Services that manage data archives and provide help identifying datasets and selecting applications, such as the NU Library Data Services, should be well marketed and coordinated to assure faculty, students, and staff are aware of them and students doing coursework or research have the necessary resources and support.
- Students need access to well-maintained public and departmental labs, especially for access to necessary software that may not be available for their personal computers.
- The campus would benefit from a central location for social science research computing, perhaps the CRESAP computing lab, that would bring together students, teaching assistants, and faculty from across the University for instruction and research, provide a place for tutorials and consulting, and establish a community that can learn from each other.
- Student instruction in research methods is typically included as part of courses or summer workshops. Establishing a common research methods course to teach students across different disciplines how to use necessary applications, work with datasets, and become comfortable with statistical methods should be explored.

RECOMMENDATIONS

The Committee reviewed a number of approaches to advancing social science research computing at Northwestern University. It was concerned with not only providing new computational and storage resources for current users of the Cell, but addressing the unmet needs of current and potential participants, including a large community of researchers using Windows systems. In addition, the Committee was interested in an approach that could provide a sustainable model for the ongoing maintenance and expansion of central as well as local resources.

The Committee recommends the University address the needs of social science research computing in the following ways:

- ***Migrate the Current DCE/DFS Cell to Intel-based Systems Running Linux***

Even though DCE/DFS for Unix has been losing commercial support over the last few years, the move by researchers, especially at national research laboratories, to Linux on lower-cost Intel systems has revived demand and support. DCE/DFS clients are now available for Redhat Linux 7.1. Because the Committee expects support for DCE/DFS on Linux will continue, it recommends social science research computing remain with the DCE/DFS environment, and use new equipment purchases of Intel-based systems to migrate the Cell to Linux.

Intel-based Linux compute servers would replace the existing two centrally purchased Unix servers. Faculty, departments, and research groups would replace and expand their own computation resources in the Cell with lower-cost Linux-based equipment, as needed. The central file server would also be replaced with Linux-based equipment to provide expanded storage capacity. Individuals and research groups could purchase additional storage for the central file server or create their own storage resources. All existing compute and file servers in the Cell could remain in service until they are replaced.

Benefits: Migrating the Cell to Intel-based hardware running Linux would result in minimal disruption of current faculty research, graduate student dissertation research, and undergraduate course work. The high level of support services for computation, storage, and applications for shared and restricted resources that faculty using the Cell currently enjoy would continue. The option of purchasing commodity-priced hardware will encourage faculty, departments, and programs to address their computation and storage needs and make it much easier to establish an annual operating budget for maintaining resources.

Support: Because the resources of the DCE/DFS Cell are managed centrally over the campus network, current staffing of 1 FTE should be able to support expanded and unmet demand for computation, applications, and storage in the Cell.

- ***Create a Linux Cluster for Computationally Intensive Research***

The Committee recommends creating a small cluster of Intel-based computers running Linux to address the needs of technically sophisticated researchers with computationally intensive tasks that are not easily address by the Cell.

The cluster would not be part of the DCE/DFS Cell. Faculty would submit jobs to the cluster but would not work interactively. Each node of the cluster would have its own memory and scratch disk. The nodes would be connected to each other by a high-speed network switch. With no permanent storage, faculty would be responsible for moving their data to and from the cluster.

Benefits: The cluster would address the needs of technically sophisticated researchers with computationally intensive tasks, such as long-term serial simulations or parallel applications. Because the cluster would be built around lower-cost commodity hardware, a budget for adding or replacing nodes each year can be more easily sustained. The cluster would provide a cost-effective model for faculty and programs interested in creating their own advanced computational resources. While the cluster would not be appropriate for general research computing initially, as cluster software continues to evolve the cluster may become easier to use, more interactive, and accessible through the Cell.

Support: NUIT Academic Technologies has been evaluating cluster technologies for the last two years and estimates .2 FTE would be required to support a social science computational cluster and assist faculty and programs interested in creating their own clusters. Experience in the Chemistry Department has demonstrated that a cluster can be easily administered, even as it expands by adding or replacing low-cost nodes.

Application support to help faculty take full advantage of the cluster architecture for parallel processing that would affect the practice of Social Science research computing would require an additional FTE.

- ***Support Windows-based Social Science Research Computing***

For technical reasons, Windows users cannot be integrated into the Cell. They can, however, remotely login to the Cell from their computer and work in the Linux/Unix environment of the Cell, or they can install Windows versions of the applications they need on their own computers and run them locally. Staying in the familiar Windows environment simplifies work, but is limited by the speed, memory, and storage of the computer.

The Committee recommends Windows users have access to network storage for datasets and research files. This can be accomplished either through a central SAMBA server or purchasing a Windows DFS client for each participating Windows computer, which would improve network performance. Anticipating network storage capacity needed may be difficult. With no campus-wide network storage available, any network storage is likely to be used to backup personal files, applications, and even MP3 files.

Benefits: Because Windows is the most common and familiar environment for students and faculty, it is playing an increasingly important role in social science research computing. While it is technically not possible for Windows users to run applications in the Cell without logging in remotely and using Unix/Linux, providing network storage would be an important new service for social science research computing.

Support: Providing support for network storage and consulting services will require .5 FTE. Extending social science computing resources for Windows users will raise a number of other important support issues. Windows users will need support for installing and using specialized applications they may not be accustomed to, setting up and using network storage, and, if needed, learning how to take advantage of the more advanced Unix/Linux-based resources available in the Cell. The costs of Windows software, preferably through site-licenses or special purchasing agreements, DFS clients (\$150/computer) when improved network performance is required, upgrading local network connections from shared to switched Ethernet for working with large datasets, and increasing central storage capacity to meet demand should be anticipated. An increased number of users will likely make the need for consulting services for statistical methods, identifying and using datasets, and using applications even more important. The Committee recommends that these, and other, support issues be further explored by NUIT and the schools and programs most in need of Windows services.

- ***New Applications: Geographic Information Systems (GIS)***

The Committee was also interested in identifying new areas of opportunity for social science research computing that are not being met by the current suite of software or would be of interest to fields that have not typically relied on computational methods

The Committee recommends adding Geographic Information Systems (GIS) software to the collection of applications available to faculty and students. GIS makes it possible to create and manipulate layers of geographically referenced information about the physical environment, man-made infrastructure, demographics, etc. to provide a better understanding of a location.

Benefits: GIS provides important new opportunities for researchers using either Unix/Linux or Windows and should particularly benefit undergraduates

Support: In general, there is need for greater applications support, whether from the NU Library for NU Information Technology. Introducing and supporting GIS would be best served by 1 FTE in NUIT Academic Technologies to provide training, classes and consulting to NU researchers and students and 1 FTE to provide new services and support in the NU Library Data Services

A budget estimate is included at the end of this report for the initial hardware/software purchases to begin the migration of the DCE/DFS Cell to Linux, creating a new social

science computational cluster, providing network storage to Windows users, and making new software available to the social science research computing community. Wherever possible, solutions are based on using lower-cost Intel systems running Linux. An estimate of annual costs to assure an ongoing 3-year replacement cycle for central resources is also provided.

CONCLUSIONS

Experience with the DCE/DFS Cell has demonstrated the value of being part of a community using a common set of tools and resources that makes it easy to share experience and expertise. It has also provided some important lessons.

- There is a significant cost when people cannot get their work done. If faculty and students do not have the right software, there is not enough storage to work with, or there is inadequate computation power to analyze data or run simulations in a reasonable amount of time, they will find inefficient, time-consuming alternatives or reduce the scope of their endeavors.
- Faculty and programs should be able to buy into the environment by adding computation or storage resources. This is especially important for heavy users who might otherwise overwhelm shared resources. However, for this to be a sustainable model the expense must be cost-effective and competitive with purchasing computation or storage for independent use.
- It is important that central resources keep up with demand and the technology be refreshed to stay up-to-date. Not maintaining and updating central resources discourages investment in a shared model and encourages faculty and programs to look elsewhere for resources and support.

The committee believes that its recommendations addresses these issues and offer the following advantages:

- Migrating the DCE/DFS Cell to lower-cost Intel systems running Linux will increase resources and capabilities for faculty and students with little disruption to ongoing research and instruction.
- Expanding the computation and storage capacity for the Cell will greatly benefit the current user community and encourage faculty and their students not currently taking advantage of the Cell's resources and capabilities to participate.
- The computational cluster will provide a new model for supporting advanced social science research computing that will become increasingly important across the University and will likely benefit the broader social science research community, over time.

- Access to network storage provides a significant new service to Windows users and begins the process of integrating them into the larger social science research computing community.
- Adding to the collection of available application software, especially GIS, will support new areas of social science research.

The Committee believes that moving to lower-cost Intel-based systems running Linux will make it possible to establish a sustainable model for maintaining performance and capacity for social science research computing. Following the Committee's recommendations will result in a more robust, contemporary environment that will attract and retain faculty and graduate students, support a broader community of researchers, and make it possible to enhance graduate and undergraduate social science education to better prepare students for research careers and advanced research.

APPENDIX A:
SOCIAL SCIENCE RESEARCH COMPUTING
AT OTHER UNIVERSITIES

The University of Chicago

Stanford University

University of Michigan – Ann Arbor

The University of Wisconsin – Madison

The University of Chicago

The University of Chicago **Social Science Division Computing Services, SSDCS** (<http://ssdcs.uchicago.edu/>) operates as a cost center, charging departments in the Division for a range of basic services. There are three support areas for the SSDCS: desktop support for faculty (Windows, Mac, Solaris, and Linux), lab and server support for four of the SSD's seven departments (the Institute for the Mind and Biology, Anthropology, Political Science, and the Center for Computational Psychology), and administrative support for the Division. This includes home pages for the division, its departments, and committees, server and database support for the Local Business Center, and desktop evaluation and replacement for administrative staff.

At least six full-time staff and over five half-time students report to the director. A faculty oversight committee provides input to SSDCS. Among the unit's services are free nightly backups for all faculty workstations. SSDCS manages about 20 servers within the Division, supporting a variety of file systems, including AFS and NIS. There are no direct charges for management of these servers, and faculty are given both storage and processing cycle quotas, with the option to purchase more space or cycles from grant income.

Unlike SSDCS, **Social Science Research Computing, SSRC** (<http://www.src.uchicago.edu/>), is not a cost center, and currently less than 10% of its service costs are charged back to departments. With twelve staff reporting a director, SSRC oversees a variety of consulting and support services that directly serve researchers and students. The administrative SSRC Board includes the dean of the Social Science Division, the dean of the Public Policy division, the Vice President for IT, and others.

The unit oversees four labs for social science computing, has a data library group, a technical group for managing servers, and a software development group of 1.5 FTE which does development primarily in the area of social agent simulation. In addition to a 22-seat general lab used primarily for statistical computing and training, SSRC operates a small lab for the population research center, a small lab for economics, and a small advanced lab, primarily for GIS processing applications. Software consulting support and training are provided in all four labs.

Data library services are handled by two FTE's, one on permanent staff and one typically supported out of grant funding. Besides providing liaison service to ICPSR and other centralized services, the University's data library currently provides services for approximately 5,000 data sets held at the University. A new "data convertibility center," transfers data from one medium to another, including floppies, CD-ROMs, Zip disks, tape, etc.. The data convertibility center can also digitize video for faculty in the social sciences.

Demands for SSRC consulting services are growing at the University. Though for many years offering primarily social statistical computing support, GIS-related services are definitely a growth area. It is apparent that a high standard of service is the norm; as an example, students can purchase statistical software at a substantial discount (see <http://licensing.uchicago.edu/list.html>.)

Stanford University

Social Sciences Data Service (SSDS), like other general information technology services, is a unit of the Stanford University Libraries and Academic Information Resources (SUL/AIR). Background information on SSDS may be found at <http://www-sul.stanford.edu/depts/ssrg/ssds/>.

SSDS provides support for faculty and student research using large datasets and a variety of statistical analysis software. All the major packages – SAS, SPSS, Stata, Minitab, and others – are available for use in the SSDS facility. There are twelve PC workstations available for users who come in to SSDS, six of which have the major utility packages and six offering a variety of software packages for specialized data analysis, such as network analysis and other methodologies. Macintosh workstations within the University Library provide key-served access to the common packages as well. The SSDS operation services the University's ICPSR membership, a Roper poll data archive, and a wide variety of statistical data from government and research agencies around the world.

Data storage for faculty research at Stanford is ample, and in most cases facilitated by the Leland data network and the declining cost of disk space available to individual social science researchers. Leland's AFS file system provides staging space for data uploaded from ICPSR or other archives, thus making the data available for researcher download to local machines.

Within the SSDS there is also a consulting service, **Statistical Software Support** (<http://www.stanford.edu/group/consult-stat/>), offering both faculty and students advice and instruction on the use of major statistical packages. This service is staffed by part-time graduate students in the social sciences who are experienced with statistical analysis.

Support of GIS software is handled through at least two service centers at Stanford, the Statistical Software Support unit mentioned above, and the Earth Sciences Library. The Earth Sciences Library oversees the University's ESRI site license (ArcView, ArcInfo, and other products), and maintains a web site on Stanford GIS resources at <http://www-sul.stanford.edu/depts/gis/gishome.html>. Nearly twenty online-based courses on the use of GIS software are available for Stanford University faculty and students at no charge, through an arrangement with ESRI.

University of Michigan

The University of Michigan's **Institute for Social Research**, or ISR (<http://www.isr.umich.edu/>) is the largest university support organization for social science computing in the US, if not the world. ISR employs over 400 regular staff and 1,000 interviewer staff, offering a range of services not only to Ann Arbor faculty, but through its **Inter-University Consortium for Political and Social Research (ICPSR)** for researchers around the world. The ISR support base includes approximately \$70 million in revenue for the services it provides.

Services to faculty in the social sciences are provided on a cost basis, and the majority of the machines they support now are PC's, with only a small number of Unix workstations in faculty offices. Providing desktop support, electronic mail support, and managing approximately 40 Unix servers comprise the bulk of their day-to-day activity. They also offer computer purchasing and setup services (at a consulting charge of \$65 per hour).

ISR is actually an overarching organization for several units, including the Survey Research Center (the largest unit), the Population Studies Center, the Center for Political Studies, and the Research Center for Group Dynamics. The ICPSR is managed out of ISR, but is a separate corporate entity from the University.

Network storage is provided for faculty in the social sciences, with a routine allocation of 100 megabytes per researcher and more available as needed. The data is backed up nightly. Note that through the University's "Basic Computing Package," or BSC, EVERY student, faculty, and staff member is provided network storage quota of 50 megabytes.

Separate from ISR is another service, under the office of the Vice President for Research at the University, called the **Center for Statistical Consultation and Research**, or CSCAR (<http://www.umich.edu/~cscar/about.html>). Though not limited to social science research support, CSCAR offers consulting support for faculty and graduate students, seminars and workshops (for a fee), and services to vend software packages such as SPSS, SAS, and others. Pricing is discounted significantly.

Free support and consulting on the use of large datasets is also provided by the University Library's **Numeric Data Services** unit (<http://www.lib.umich.edu/data/index.html>). With a staff of three, including a data librarian, a statistical consultant, and a programmer/analyst, the service of this unit primarily supports users of ICPSR datasets.

Beginning in 2000, the Vice President for Research has also undertaken a special two-year project for developing GIS expertise and resources within the university, dubbed the **Spatial Analysis and Geographic Information Science (SA/GIS) Initiative** (<http://gis.umich.edu/about.html>). Seven faculty projects have been funded, some as demonstration projects and others specifically to develop instructional resources for GIS use.

The University of Wisconsin - Madison

Support for social science computing at the University of Wisconsin - Madison is quite decentralized, and relies significantly on resources from the departmental or school level rather than the University as a whole.

The **Social Science Computing Cooperative**, or SSC (<http://www.ssc.wisc.edu/home.htm>), was organized as a result of federal grants to the University for support of their US Demography Center. As its name implies, the SSC is a cooperative funded by seven University units: the Center for Demography and Ecology, the Institute on Aging, the Institute for Research on Poverty, the Rural Sociology and Sociology departments, the Social Systems Research Institute, and the Wisconsin Center for Educational Research. Governance for SSC is through a board of the chairs of the eight above-named units that meets once a year. At their annual meeting, board members bring to the table the amount of funding each is willing to put in for support of the Cooperative. Besides these hosting units, a number of smaller University entities are supported through services of SSC: the Applied Population Laboratory, the Center for Demography of Health and Aging, the Center on Wisconsin Strategy, the Havens Center, the National Survey of Families and Households, the Study of American Families, the University of Wisconsin Survey Center, and the Wisconsin Longitudinal Study.

In addition to the director who serves on an annual appointment basis from faculty ranks, the SSC has eight full-time permanent staff: a statistical consultant, a help desk manager, a general administrative support person, two workstation (PC) support staff, a network administrator, and server administrators for their Windows and Unix servers. The computing resources of SSC are in labs totaling about 350 PC's, 15 Macs, and approximately 100 Winterm stations. Besides about seven independent Linux servers (SSC has been working with Linux since 2000), they administer a Condor cluster. Training provided by SSC staff is fairly extensive, with 700 to 800 users attending workshops annually, on topics such as the use of SAS, of Stata, and of common office and Unix applications.

Wisconsin's **Data and Computation Center**, or DACC (<http://www.dacc.wisc.edu/>), is a service unit of the College of Letters and Science's Social Science Division. DACC provides a Data and Program Library Service (DPLS) with a little over two FTE staff, and a Social Science Computing Lab (SSCL) with 42 networked Windows workstations. DACC's central administrative staff numbers two, beyond the director who is appointed from faculty ranks. The SSCL provides accounts for use of its lab machines to both undergraduate and graduate students, and the bulk of its user base is from the economics and sociology departments. This lab provides access to a wide range of statistical software (Minitab, SPSS, SAS, RATS, Stata, Lisrel, etc.).

GIS software resources are currently managed for the University as a whole through the **Land Information and Computer Graphics Facility**, an agency of the College of Agricultural and Life Sciences. LICGF is the distribution point for a variety of ESRI software (ArcView, ArcInfo, etc.) available at no charge to the campus community through a UW System license arrangement with ESRI.