

---

## RESOURCES

---

### University of Wisconsin-Madison

#### Social Science Research Services

Affiliates of the Center for Demography and Ecology and the Center for Demography of Health and Aging have full access to the resources of Social Science Research Services (SSRS). SSRS serves social science researchers in the College of Letters and Science at the University of Wisconsin-Madison. The central function of SSRS is to provide specialized, high quality services to advance interdisciplinary social science research at the UW-Madison. SSRS provides an administrative "umbrella" to coordinate a variety of service-providing units.

#### OFFICE RESOURCES

##### SSRS Administration (Pre- and Post-Award)

SSRS Administration maintains responsibility for overall Center management, budgeting, purchasing, editorial, and secretarial services. The Administrative staff coordinate visits of scholars from other institutions. Staff are also responsible for producing the Center for Demography and Ecology Working Paper series.

The primary goals of SSRS are to ease administrative burdens on the part of researchers involved in obtaining extramural research funding, to provide research tools and expertise necessary to carry out the research, and to assist researchers in managing their grant funds. SSRS Administration services include preparation and submission of grant applications, negotiation and liaison with units internal and external to the UW-Madison in connection to grant awards, and post-award financial management for funded research.

#### COMPUTING RESOURCES

##### SSRS Computing (SSCC)

The Social Science Computing Cooperative (SSCC) provides a full range of computing services for social science research at UW-Madison. The SSCC combines powerful computing and statistical software resources with a professional staff. Services include a help desk, statistical computing consulting, PC repair, a training program, and a large collection of on-line publications on a variety of computing topics. Instructional computing support is also provided for the faculty, staff, and students in the Social Science division of the College of Letters and Science including a classroom, drop-in lab, and mobile lab with a rich variety of statistical software. The SSCC serves approximately 1400 research members and 2000 instructional users.

The SSCC is supported by several research agencies in the social sciences at the University of Wisconsin-Madison and their individual members. Components of SSCC's costs are paid directly by member organizations and their individual members. There is no fee-for-service charge. Thus, the SSCC is the creature of its sustaining member organizations. The Chairs and Directors of those organizations meet annually to review the expenses of the previous year, to consider the SSCC's expenditure plan for the coming year and to undertake their part of its support. Policy guidance and oversight of the SSCC is provided by a Steering Committee made up of faculty representatives of the sustaining member organizations.

The SSCC operates a network in which two operating systems (LINUX and Windows) provide a high level of service to its members and instructional users. The local area network in the Sewell Social Sciences building comprises approximately 550 desktop computers and about 50 graphics terminals.

Altogether, the SSCC's servers, both Linux and Windows, host approximately 35 terabytes of disk storage on a secure RAID-6 iSCSI Storage Area Network (SAN). Backups run daily and are written to disk and LTO-tapes which are stored at an off-campus storage center.

The Linux network comprises 30 centrally-managed Linux systems, some of which are reserved for interactive logins by SSCC members and some of which support file, e-mail, printing, Internet, and other administrative services. The four most powerful servers run 64-bit Linux and support statistical calculations.

A 64-bit Linux Computing Cluster provides a significant increase in the computing power available to SSCC users. The cluster is currently made up of 72 CPU's running RedHat Linux. This cluster has a powerful batch pooling utility installed called HTCondor which was developed at UW-Madison's Computer Science Department. Besides being a powerful batch management and scheduling utility, HTCondor also provides high throughput by amassing the collective CPU cycles from the Linux computers in SSCC's cluster as well as available CPUs from other HTCondor clusters (known as flocks) on campus and even around the world.

A Linux Beowulf Cluster provides researchers with the ability to run parallelized code written in FORTRAN and C/C++ for estimating and solving large dynamic programming models that can be "broken" into pieces (e.g., spatial models, labor market equilibrium, large estimation projects, etc). The cluster is comprised of over 200 cores.

The Windows 2008 network comprises 40 centrally-managed Windows systems. The domain provides network access to PC home directories and other shared disk space, a large number of printers, important Internet services, and some network software. Many of SSCC's network and Windows administrative services have been virtualized in recent years, resulting in a 90% reduction in the number of physical administrative servers.

The 22 most powerful servers run Windows Server 2008 R2 Remote Desktop Services, a multi-user adaptation of the Windows operating system by which SSCC members can run Windows statistical software and other Windows programs from any location. Linux is still the preferred platform for running the most CPU-intensive jobs, but with each Remote Desktop Server's two 8-Core Intel Xeon processors (for a total of sixteen physical processors) running at 2.6 GHz and 64 gigabytes of RAM, it is a robust and viable platform for members running large statistical jobs. These Remote Desktop Servers are enormously popular with SSCC members and are extremely effective both in reducing hardware and software acquisition costs and in centralizing the administration of Windows software and PC hardware (and thereby lowering the support costs).

Co-op users spend most of their computing time on three categories of tasks: statistical analysis of data, text editing and document production, and electronic mail. The most commonly used applications for statistical analysis are Stata and SAS. Other statistical applications available on the network include aML, Amos, ATLAS.ti, Gauss, HLM, Lisrel, Maple, Mathematica, Matlab, MIwiN, MPlus, Nvivo, R, and SPSS. Non-statistical applications include ArcView, EndNote, LaTeX, Microsoft Office, Adobe Design suite, Scientific Word, and SmartDraw.

## **OTHER RESOURCES**

### **Data & Information Services Center (DISC)**

SSRS provides data services through the Data and Information Services Center (DISC). DISC services include reference, data acquisition, cataloging and classification of datasets, aid in acquiring restricted data sets, preservation of data, making public use data available via value added data extraction systems, data current awareness services, and campus outreach services to promote the use of social science data to the larger campus community. DISC librarians are available as guest speakers to academic classes, providing instruction on social science data resources and practices.

DISC maintains one of the world's finest research collections of machine-readable data files in demography. The collection of over 3300 data files is available on Unix disk (backed up regularly on 8mm tape), CD-ROMS, and tape cartridges. The collection is strongest in U. S. census data and large household surveys; fertility, vital statistics, life history, and social mobility surveys. In particular, the collection is strong in the areas of U.S. decennial census files, Current Population Survey files, National Health Interview Files, and Mortality and Natality Detail files.

Many data files in the collection are more than 100 megabytes in size; the largest is 4 gigabytes. Data are available in compressed and uncompressed formats. Increasing numbers of data sets are acquired on CD-ROM accompanied by a diverse assortment of extraction and retrieval software. The more than 50 extraction programs are installed on a PC with temporary storage space for users to create data extracts.

As the availability of demographic data resources via the Internet continues to grow, and as new methods for storing electronic data make old storage media obsolete, preservation of valuable data files is an ongoing concern. To that end, DISC preserves its data on optical media (CD-ROM), while at the same time constantly following the digital preservation literature to see what new formats and media are in development. At present, data is preserved on over 320 CD-ROMS.

One primary mission of DISC is collection, quality control checking, organization, and preservation of data sets for the use of social science researchers. The DISC merger added about 1600 datasets from the old Data and Program Library Service to the merged collection. The "Ready Reference" data set collection, data sets from which researchers can quickly extract descriptive data for research or training purposes, includes subjects such as world population, welfare, health, mortality, and macro-economic data, among others. There are 27 datasets in this collection. In addition, DISC maintains a pointers page to Country statistical yearbooks from 152 countries and NGO agencies (<http://researchguides.library.wisc.edu/content.php?pid=24954>)

DISC's World Wide Web online catalog has grown as new datasets have been added. The datasets are cataloged, and most can be accessed via key word search. DISC staff maintain and update Web pages that provide annotated links to over 1,000 data-related resources on the Internet. (<http://www.disc.wisc.edu/newcrossroads/index.asp>). These links make it easier for researchers to access U.S. and international population data, microdata files, longitudinal studies, vital statistics, health, and macro-economic data. DISC also maintains two current awareness services for CDHA (the Current Awareness in Aging Research weblog-- <http://www.ssc.wisc.edu/cdha/CAAR.html> and the daily Current Awareness in Aging Research E-Clippings—available only in e-mail due to the unstable nature of newspaper Internet addresses). These newsletters have a subscription base of around 3,500 worldwide. In addition, DISC maintains a weblog of new Internet resources for the Center for Demography and Ecology (Current Social Science Research Reports—(<http://www.disc.wisc.edu/reports/cssrindex.html>)).

## **Center for Demography of Health and Aging**

The Center for Demography of Health and Aging (CDHA) complements the Center for Demography and Ecology (CDE) at the University of Wisconsin-Madison, an NICHD-supported R24 center for demographic research. CDHA both shares and builds on the substantial research infrastructure of CDE. It creates links between social demography and biomedical and epidemiological research on aging. The overall goal of the CDHA is to advance a major research and training program in the demography of health and aging.

Major themes of ongoing and developmental research activities within CDHA include (a) midlife development and aging; (b) economics of population aging; (c) inequalities in health and aging; and (d) international comparative studies of population aging. These cross-cutting themes build on current activities of members of CDHA.

CDHA activities revolve around five core areas:

- § The administrative and research support core provides leadership and administrative support for the Center.
- § The program development core provides modest support for new faculty development and for faculty, staff, and research assistants engaged in innovative pilot research projects that are likely to lead to major NIA support.
- § The external innovative network core provides a networked current awareness service for research in the demography of aging and supports regular workshops, conferences, and visits.
- § The external research resources support and dissemination core supports user-friendly provision of large-scale public data resources in the demography of health and aging.

§ The statistical data enclave core, which encompasses the organization and archiving of data collections carried out by CDHA, as well as offering data protection for restricted datasets used by CDHA researchers.

### **Data Protection at SSCC/CDHA**

The purpose of this security protocol is to protect the confidentiality and integrity of sensitive data. For these purposes, sensitive data are defined to include any information defined as “sensitive” by the Computer Security Act of 1987<sup>1</sup> and proprietary research data. Preserving confidentiality consists in ensuring that sensitive information is not accessed by unauthorized people. Preserving integrity requires ensuring that information is not altered by unauthorized people in a way that is not detectable by authorized users. The Center for Demography of Health and Aging (CDHA) in cooperation with the Social Science Computing Cooperative (SSCC) at the University of Wisconsin – Madison (UW) offers four different levels of protection in an effort to satisfy the requirements of data producers and the needs of researchers who wish to analyze these data. These four levels, listed in order of increasing levels of security, are terms: Protected, Encrypted, Restricted Access, and Remote Access.

Identified threats to confidentiality and integrity of sensitive data are as follows:

1. Internal failure (e.g., hardware failure, software bugs)
2. Environmental (e.g., fire, flood)
3. Personnel (e.g., prolonged illness)
4. Outsiders (e.g., theft or disclosure of confidential information).

Security-related activities have been designed to minimize the impact of these threats thereby ensuring data confidentiality and integrity. These activities can be grouped into three categories: prevention, detection (when prevention fails), and recovery. The compensatory activities promote redundancy (ensuring that the system operates despite damage to an individual part) and defense-in-depth (requiring that an attacker overcome a series of layered barriers before accessing or damaging the system).

### **Level 1: Protected Access**

#### **Physical Security**

The SSCC Windows and Linux servers are stored in a locked humidity and temperature controlled room. Physical access is restricted to SSCC computer staff and entry is controlled with a card reader.

#### **Account Security**

Account Security consists of a username-password combination. Accounts are set up so that only authorized users have read, write, and execute privileges. Other users and outsiders are not assigned read or write privileges, for obvious reasons. As an added layer of protection against hacking, CDHA recommends setting up default privileges so that only users who require access to the directory have execute privileges. With the exception of the system administrator and his staff who maintain the system, users are not assigned sudo privileges, meaning that only administrative staff and the authorized user(s) have access to the data. CDHA recommends locking the computer screen or logging off whenever leaving the workstation. CDHA also recommends against keeping sensitive data on local hard drives.

#### **File System Security**

Both the Windows and Linux networks are protected by a firewall. The firewall ensures that only authorized

---

<sup>1</sup> The definition of sensitive information contained in the Computer Security Act (1987) is as follows: *any information, the loss, misuse, or unauthorized access to or modification of which could adversely affect the national interest or the conduct of federal programs, or the privacy to which individuals are entitled under section 552a of title 5, United States Code (the Privacy Act), but which has not be specifically authorized under criteria established by an Executive Order or an Act of Congress to be kept secret in the interest of national defense or foreign policy.*

users within the Social Sciences Building can get to data stored on the network.<sup>2</sup> File system security consists of monitoring and auditing system log files by the system administrators. This allows for the systematic evaluation of system activity, system malfunctions, and security breaches. In the Linux system, for example, the system administrator monitors repeated failed attempts to log in as the system administrator, use of the sudo command which assigns administrator privileges to the non-administrative user, and unauthorized shutdowns that can be used to modify log files to cover an intruder's tracks. Redundancy is accomplished via nightly backups to tape. Backups are stored in a secure humidity and temperature controlled fire proof off-site location.

### ***Level 2: Encrypted***

In addition to the protections described above, data files may be encrypted for extra security. CDHA recommends that users analyze data using SAS and encrypt data using the SAS Security software because this software processes decrypted data in memory only, which means that unencrypted data are never written to disk. If users want to use STATA or other software to analyze data, it is possible to encrypt the data using PGP or other reliable encryption software. In selecting encryption software, we recommend choosing one that uses the new AES algorithm (Rijndael Algorithm).<sup>3</sup> If you use encryption software other than SAS Security, it is important to pipe the decrypted data directly into the analysis software and to ensure that memory limitations are not exceeded in order to prevent writing decrypted data to disk.

### ***Level 3: Restricted Access***

SSCC has created a secure cooperative computing facility which we call the "cold room." The cold room contains a single stand-alone workstation and can be accessed by only one researcher at a time. The operating system and software are stored on a permanent (non-removable) hard drive. Data are stored on a removable drive that is accessible only to individuals with licenses for sensitive data.

To gain access to the cold room contained within this outer room, the researcher must have a key. Approved IRB protocols are issued keys for the cold room. As an extra security measure, the lock on the cold room is keyed to a security master, meaning that it is not accessible using the master keys carried by custodians or other university faculty and staff. The Director of DISC is the only person authorized to issue keys to researchers who wish to use the cold room.

Each researcher authorized to use the cold room is assigned a removable hard drive (the "data drive"). This data drive is stored in a safe assigned only to the licenses data user. The safe is located within the cold room. This person may assign an 8-digit security code to the safe if required by her data license. When the licensed data user enters the unoccupied computing area, s/he opens the safe and inserts the data drive into the computer, boots the machine and enters a password. When the researcher is ready to leave the cold room, the computer is shut down and the data drive removed and stored in the wall safe. The operating system automatically starts up a password protected screen saver when the keyboard and mouse are idle for 5 minutes. This ensures that only the licensed data user has access to data files.

---

2 There are two exceptions: (1) authorized SSCC users can access areas where data is stored using VPN, and (2) the web servers are accessible to outsiders.

3 Additional information on federal encryption standards is contained in the Cryptographic Toolkit (<http://csrc.nist.gov/CryptoToolkit/aes/>) published by the National Institute of Standards and Technology. Updated information is also available on the NIST website ([http://csrc.nist.gov/focus\\_areas.html#csa](http://csrc.nist.gov/focus_areas.html#csa)).