# A Research Cooperative for Social Scientific Computing: Development, Status, and Prospects[1]

**Robert M. Hauser**
**Vilas Research Professor of Sociology**

Center for Demography and Ecology
The University of Wisconsin-Madison
1180 Observatory Drive
Madison, Wisconsin 53706

CDE Working Paper 87-41

December 1, 1987

# A Research Cooperative for Social Scientific Computing:

## Development, Status, and Prospects[1]

**Robert M. Hauser**
**Vilas Research Professor of Sociology**

Center for Demography and Ecology
The University of Wisconsin-Madison
1180 Observatory Drive
Madison, Wisconsin 53706

CDE Working Paper 87-41

December 1, 1987

---

[1] Remarks prepared for a conference, "Innovations in Social Science Computing," at the Cornell Institute for Social and Economic Research, Cornell University, Ithaca, New York November 11-13, 1987. This is a revision and update of an essay originally prepared for the Max Planck Institute for the Study of Education and Human Development in Berlin and published in Bericht der EDV Kommission, Anlage 2, 1985. I thank Bruce Ruzicka for helpful comments on an earlier draft of this paper. The Center for Demography at the University of Wisconsin-Madison receives support for population research and training from the National Institute for Child Health and Human Development and from the William and Flora Hewlett Foundation.

To provide an appropriate context for my remarks, let me begin by describing the faculty computing cooperative that I manage. Then I will describe how it grew, how it works, and some of its major problems and future prospects.

The CDE/SOC/AGE co-op is a coalition of several university units--The Center for Demography and Ecology, the Departments of Sociology and of Rural Sociology, and the Institute on Aging and Adult Life--the mission of which is to supply computing services to its members. All of the units served by the CDE/SOC/AGE co-op conduct significant research in social science and health related fields. Although the co-op serves a large and somewhat diverse group of researchers, it is not a general-purpose or full-service computing facility, it deliberately excludes several major components of social scientific computing at Madison, and its leadership and members have no desire to see it play a larger role in academic computing.

Located in the Social Science Building, the CDE/SOC/AGE co-op manages three VAX computers--a VAX 11/780 (owned by the Center for Demography and Ecology), a second VAX 11/780 (owned by the Department of Sociology), and a MicroVAX (owned by the Institute on Aging and Adult Life). In addition to the three VAX computers, the facility includes 13 disk drives, 6 tape drives, 6 large printers, three networks, 200 office connections and terminals, dozens of PC's and small printers, data base management systems, statistical and analytical software, compilers, utilities, and custom software.

Earlier this year, the CDE/SOC/AGE co-op assumed management responsibility or the Social Science Building's high speed Ethernet computer network. Currently the Institute for Research on Poverty's MicroVAX computer is attached to the Ethernet, and there is capacity for further large computer connections by others in the building.

The CDE/SOC/AGE co-op is staffed primarily by persons employed by the Center for Demography and Ecology and (to a lesser extent) by the Department of Sociology and the Institute on Aging and Adult Life. The Data Processing staff consists of professional academic, classified, and student hourly workers that support, in varying degrees, the research activities of the faculty. There are approximately ten academic, one classified, and eight student hourly staff who operate and support the computer systems and networks.

The systems are in operation 24 hours per day, allowing simultaneous interactive use by up to 150 of the nearly 500 unique users. These facilities are presently connected to computers on the campus and elsewhere through the campus high speed Broadband networks. These computers include the VAX clusters at the Physical Sciences Laboratory and the Madison Academic Computing Center. Efficient operation of the system is extremely important because the CDE/SOC/AGE co-op does not have a resource charge-back system to assist in allocation of resources; the system relies on a level of knowledge and spirit of goodwill among its users sufficient to assure its continuance.

The Center for Demography and Ecology (L&S) is organized as a research cooperative with the directorship an elected and rotating position among senior faculty. I currently serve as Director. The staff of the Center are organized around four core units: Data Processing, Data Library, Print Library, and Administration. Center staff do not participate directly in research projects, but provide the resources necessary for scholars to pursue their individual research interests. Among those resources are a networked VAX 11/780 computer, extensive libraries of both censuses, large scale survey data, and other printed material, and access to administrative support services.

The Department of Sociology (L&S) and the Department of Rural Sociology (CALS) participate in and contribute to the support of the CDE/SOC/AGE computing co-op. All faculty members in both departments are entitled to use the SOC VAX 11/780 for research and graduate teaching. A Sociology Computing Committee, currently chaired by Professor H. Andrew Michener, consists of faculty members from the two departments as well as computing staff and representatives from CDE and AGE. This committee sets policy for the management and operation of the SOC VAX 11/780. Funding for the ongoing costs of this facility come from four basic sources: the College of Letters and Science, the College of Agricultural and Life Sciences, the Graduate School and individual research grants of faculty and research staff using the resources.

The Institute on Aging and Adult Life is directed by David L. Featherman, John Bascom Professor of Sociology. The Institute is a multidisciplinary research and graduate training unit within the Graduate School of the University. The Institute's many activities including coordinating and encouraging the development of research and curricula throughout the campus and assisting researchers through funding and information resources.

**Some Caveats**

Before I go too far in outlining my ideas about the state of the art in social scientific computing, I think it important that I say a few words about my qualifications to give advice about computing. I offer no advice as a technical expert in computing matters. I am not a computer scientist, systems analyst, or programmer. I wrote a few lines of FORTRAN while in my first year of graduate study 20 years ago and decided that I would rather become a sociologist than a computer programmer. I have been a very heavy user of statistical computing facilities over the years, but until about 6 years ago I always worked through a layer of research assistants and programmers. I still do no programming as such, although I now spend my work days (and nights) in front of a video terminal or a micro-computer instead of behind a pencil and writing pad.

I do not outline these limits of my expertise with any apologies - there are lots of technicians around when one needs answers to technical questions. My expertise, such as it is, comes from having observed and participated in the growth and development of a demographic and sociological research center at the University of Wisconsin-Madison. My main interest is doing my own work - research, writing, and teaching - up to the highest possible standards and as quickly and efficiently as possible in real time. That has sometimes meant avoiding the well-intentioned enthusiasms of people around me who would rather spend time trying out the latest in high technology than in getting a job done. Such enthusiasms are all too prevalent among computing experts (and some non-experts as well), and this is one of the things that have led me to conclude that computing is too important to be left only to the experts. The design and management of a computing environment cannot be successful without continual guidance from users of the facility, including those with no specific expertise in computing. Moreover, many of the most important issues are not technical at all, but pertain to questions of resource acquisition and allocation, both in the long and short run.

**Computing at Madison**

During the 1960s and early 1970s, William H. Sewell and Halliman H. Winsborough were successful in bringing a highly productive group of young sociologists to Madison. Many, but not all of these scholars carried out quantitative research, and several - especially a group recruited from the University of Chicago and the University of Michigan - had been trained in demography. The members of the Center for Demography and Ecology (CDE), myself included, formed a little research cooperative around a small core of money provided by a federal grant for graduate training and the collection of (printed) library material. We pooled these resources and whatever other research support we managed to bring in from the outside, operating on the principal that the creation of a common resource base - our catch-phrase was "critical mass" - would ultimately bring in more resources for everyone than would the usual model (in the U.S. at least) of individual research entrepreneurship. Resource-pooling has persisted during the growth of the CDE to include about 250 regular users of our core facilities, plus another 350 or so users of related facilities under CDE management, and it is an important feature of the organization of the Center. I think this point is very important in illustrating that many of the problems of designing a productive computing environment are in fact problems of social organization.[2]

By the early 1970's we were specializing in the analysis of large and complex data files from surveys and censuses, for example, the 1/100 samples from the 1960 and 1970 U.S. Censuses and samples from the monthly Current Population Surveys. The main University Computing Center (MACC) was not well equipped to manage such large tape files, and its cost structure was much friendlier to heavy number-crunching than to the management and analysis of large bodies of data. Another problem with MACC, which has continued to plague the organization to the point where it has almost ceased to exist, has been the commitment of the University of Wisconsin to decentralized, pay-as-you-crunch funding. The proliferation of departmental computing, the development of a competitive off-campus

---

[2]In the story that follows, I have mainly observed from the sidelines. Most of the credit for the success of the CDE belongs to Hal Winsborough, Peter Dickinson, and a succession of Center Directors - James Sweet, Larry Bumpass, and Karl Taeuber, as well as to our Directors of Data Processing, most recent among whom is William Gates, and to many other able staff members.

VAX shop at the University's synchrotron, and the infiltration of PCs have virtually driven MACC out of business. At present, its main, shared facility is a VAX cluster, built around an 8600 that typically supports about half as many interactive processes as our cooperative, and its users are burdened by a cost structure that attempts to support the amortization of equipment at rates that ignore the secular decline in its cost. Finally, frequent and unpredictable changes in MACC's system had major effects on our ability to work. At that time, MACC had a monopoly on the provision of computing services through its UNIVAC mainframe, and we had to negotiate for some period to convince the University administration that our specialized needs could not be met at reasonable cost using MACC facilities.

Ultimately, we received permission to rent a small IBM 360 computer with three tape drives, a card-reader, card-punch, and line-printer on the condition that we would not compete with MACC in general purpose computing. That is, we were expected to manage and summarize data on our machine, but to hand-carry data to MACC on cards or tape for further statistical analysis.

As time went on, MACC supervision of our work declined, other competitors to MACC sprang up around the campus, and we carried out an increasing amount of analytic work on this little machine. For example, the hundreds of log-linear models that David Featherman and I ran in connection with the 1973 Occupational Changes in a Generation project were run for long periods, late at night on that machine. Even with the generous support that our project had from the U.S. National Science Foundation, we would never have been able to pay for all that analysis on the main University computer. Occasionally, we brought new software to the campus. For example, we introduced SPSS to Madison, and we had to buy the program for MACC in order to keep potential users away from our machine.

In keeping with our position on sharing resources, we ran this machine on an "open shop" basis, initially without paid operators. That is, each faculty member, staff member, or student assistant in the Center was entitled and encouraged to take a few hours of instruction in the operation of the computer and then to sign up a day or two in advance for whatever block of time seemed necessary to do his or her work. (I never went near this machine.) As time went on, we found that it would be fairer and more efficient to hire a student operator to run short jobs for people during regular working hours, and to reserve "open shop" hours for the late afternoons, evenings, and weekends.

There was support for one or two programming positions and a data librarian in the core grant, but mostly we developed software to solve common problems through programmers who were hired on individual projects. This was both defensible and efficient, for we often did similar things with different bodies of data. This early commitment to staff development of common resources has persisted: There has been little work by central staff on individual research projects; obversely, project programmers have often contributed shared software resources to the commons. We have tried to maintain an environment where there is minimal need for custom programming and where it may even be slightly disrespectable. Although we have, from time to time, taken on new groups of users, as well as fresh cohorts of graduate students, the central idea of the place has never been to provide general support for computing; rather, we are a research shop for people with common resource needs who know what they are doing. We have the expectation that most users will become capable of doing most of their work for themselves; we have the continuing problem of persuading new users to believe, first, that they can learn to do their own work and, second, that no one else will do it for them (or does it for anyone else).

Throughout the years, in obtaining new hardware and software, we have tried to find a place just behind the latest technological frontier, and to stay there for some time before moving onward. That is, we have tried not to bear the burden of developing or testing new technologies, and we have also tried to create a working environment that would be stable for long periods of time. Sometime in 1975, we replaced the IBM 360 with a larger and faster IBM 370 Model 35, again on a rental basis; so far as most users were concerned, the transition took only a day, and the only changes we noticed were improvements in system performance.

**The CDE VAX**

Late in 1978 or early in 1979 we thought that it was time for a major change in the system. IBM was introducing a major new line of minicomputers that appeared to make interactive, time-sharing computing feasible for the first time, but there were also several competitive minicomputers. Moreover, the Center grant was up for renewal, and we were concerned about the continuity of funding beyond the forthcoming five-year renewal period. What if a new Federal administration or Congress were to decide that demography should no longer be supported? What resources would remain for us?

After some investigation, we found that the outright purchase of a VAX 11/780 from Digital Equipment Corporation would be fully competitive with rental of a new IBM machine, and we were able to persuade NICHD that it would be a good investment. The VAX (virtual memory) technology was a couple of years old. It had proved successful in many applications, and the VAX outperformed several other competitors at the types of input-output operations that were critical to our use of large data files. With this machine, several people could work with the computer at the same time from any location - programming, writing, organizing the information in their personal files, executing or submitting programs, reviewing output from programs, or ordering output to be printed - and in addition a relatively unskilled operator could mount and dismount tapes as needed and operate the line-printer. The main problems at the time were that the VAX was in such high demand that there was a wait of several months before we could get one, and the purchase order was so large - about $350,000 - that we had to have the signature of the Governor of Wisconsin to proceed.

When our VAX was installed in the fall of 1979, it had 1.5 megabytes of memory, 16 ports for interactive computing, 225 megabytes of hard disk storage, and a line printer. We chose the VMS operating system; its major competitor, UNIX, appeared to be better suited for business than for scientific applications. (We now think the choice of VMS may have been an error, for VMS remains proprietary with DEC, while UNIX is available from many vendors. It is difficult for us to shop for new equipment when our choices are to remain a captive DEC shop, to initiate a major system change affecting some 600 users, or to start mixing operating systems.) In any event, with the CDE VAX, we went from three down to two 800/1600 BPI tape drives - on the argument that the greater availability of hard disk storage would reduce the volume of tape jobs; this decision turned out to be a mistake. In the course of the changeover to interactive computing, we eliminated keypunch machines. We maintained a card-reader and card-punch attached to the VAX, but we soon found them unnecessary.

The transition from the batch-oriented IBM machine to the VAX was not so smooth as earlier changes in the system. There were a number of serious problems: (1) Few members of the Center had any experience with interactive computing; (2) Much of our specialized software had to be rewritten, and some standard statistical software was not yet available for the VAX; (3) The extended capabilities of the VAX changed the character and volume of our uses of the computer, and these changes, in turn, have led to continuing adjustments of the computing environment; (4) As our IBM system had aged, we had fewer problems with requests for use from the margins of demography, but the VAX was a far more attractive resource. Let me comment on each of these points with particular reference to the current development of the system.

**Group Structure and Responsibility**

The fundamental problem of Center organization and management is how to organize and allocate access to a common, finite set of resources without a price system. We organize access to the CDE VAX by analogy to Center membership. Each member of the Center Steering Committee is assigned a group number, and within that group number, he or she can assign user numbers to staff or student members of the group. Permanent disk space is assigned to each group in proportion to the number of its members, in order to permit each user to have someplace - a working directory - in which to store current working materials: mail messages, text, programs, data, or output. The permanent allocation or "soft limit" is about a megabyte per user. Project leaders are permitted to reallocate disk space among group members, and they are responsible for the proper use of the system by members of their group. Any group member can increase his or her disk allocation up to a "hard limit" to deal with temporary data storage problems within an overall group allocation, and longer-term reallocations can be made by prearrangement with the operating staff and faculty managers of the Center. If one's working materials become too large, it is straightforward to copy them to magnetic tape - or, in the case of smaller files, to floppy disks or other PC-based storage media - and to restore them to the working directory as needed. When a user exceeds the hard limit, it is no longer possible to write to disk; when the soft limit is exceeded, a time delay is imposed on subsequent logins, and if the violation is not soon corrected, the account is frozen.

Beginning in 1984, we introduced a project accounting system for use of computing resources. A group leader has one account under his or her own initials for miscellaneous unsupported work, including non-supported student research, and as many other project accounts as needed to describe funded research projects that have been approved for use of the facility by NICHD. NICHD-funded projects are automatically approved, and others are reviewed periodically. Although projects are usually subdivisions of group activities, it is also possible for collaborative projects to cut across group boundaries. Users can set default projects for interactive and batch activity, and they can easily change accounts in the course of a working session. The accounting system provides monthly and quarterly reports of use by persons, groups, and projects; we use these data to monitor the volume of funded work for NICHD, to identify patterns of activity that may suggest inefficient use of the system, and to identify areas where we may need to tune the system or to obtain additional resources.

Few members of the Center had any experience with interactive computing when the CDE VAX arrived; moreover, many of the older faculty and staff members - like me - had no hands-on experience with the earlier machines. Despite the expert knowledge of a few programmers and graduate students, the Center faced a major task of socialization. Finally, beyond our ignorance and discomfort with it, the new technology posed a structural barrier to the learning process. In the days of batch processing, it was necessary for users to congregate around the machine-room. This provided a natural setting for the exchange of information and ideas, and it was one of the most important mechanisms for the development of expertise throughout the Center. With the new technology, anyone could work wherever we could run a wire to a port - presumably each at his or her own desk. In fact, people were soon able to work wherever there was a telephone; I will say more of this later, for the dispersion of access to the machine was the source of major organizational challenges.

With the problem of socialization foremost in our minds, we decided at first to limit the distribution of computing throughout the Center - which was in fact not "centered" at all, but scattered throughout six floors of a large office/classroom building. We put almost all of the available ports in one large office, immediately adjacent to the main computer. This arrangement attracted few new users, but it did lead to rapid retraining of those who had used the old system. People congregated in the terminal room at all hours of the day and night, playing with every known feature of the system, occasionally discovering new ones, and always letting others around them know what they had done. I have used the term "playing" deliberately here. Just as in the case of childrens' learning, "play" is a very useful device for getting people to sit down at a terminal and teaching them how to use it. The VAX came equipped with a number of games - STARTREK, ADVENTURE, and HEARTS, among others - and for some months we did nothing to discourage game-playing, even during regular working hours. Now, these "utilities" are - as far as I know - no longer available. This is not an unmixed blessing, for one of our recurring problems has been the socialization of new cohorts of graduate students and other new constituencies. In any event, after a core of competent users had developed, we gradually phased out some of the terminals in the common room and began to distribute ports to individual projects and users scattered throughout the building and the campus.

Two features of the new system proved very helpful in the socialization process. First, the operating system includes a tree-structured set of documentation. Users may ask questions of the system as they work, merely by typing "HELP," followed by the name of "whatever-it-is" they do not understand, and (sometimes) the right answer will appear. When a user can not pose specific questions, "HELP" by itself produces a list of the "highest level" topics covered by the on-line documentation. The HELP facility is not limited to the software supplied with the operating system, and we have supplemented it with on-line documentation for many of the utilities and programs that we have developed or added to our system. Thus, for many purposes, it is not necessary to consult printed documentation.

Again, this is not an unmixed blessing. Some users forget that it is ever necessary to consult printed documentation, and many more find that their cognitive organization never quite matches that of the HELP files. Consequently, in addition to providing distributed libraries of frequently used manuals and a more extensive central library, we have developed other mechanisms to support users, short - I hope - of our becoming a full-service computing center. These include a continuing series of 2 to 5 page handouts on frequently used procedures and a consulting office that is open about 35 hours per week.

Second, the electronic MAIL facility made it easy to share new discoveries or to ask questions of other users or staff members. Over the past couple of years, we have developed a shared "CONSULTANT" account, accessed only by electronic MAIL, that has become the local equivalent of Ann Landers.

The MAIL facility has become the standard medium of communication among members of the Center; we use a lot less paper for routine messages. In fact some members of the Center use the VAX primarily for communication within the Center and leave the "real" computing for others to do. Most announcements of meetings, seminars, and the like circulate only by electronic MAIL, and this has become a serious problem for new and peripheral members of our user community, as well as some colleagues who have, for one reason or another, never become regular users of the system. Initially, our machine was not networked with any others, but a few of members obtained limited Center memberships for colleagues in other institutions and were so able to exchange messages, data, and work in progress. During the past couple of years, such arrangements have been largely superceded by our membership in BITNET and, within the past year, by our connection to an ethernet channel on the local campus broadband. The MAIL facility has some disadvantages. The possibility of instantaneous response invites instantaneous response, and many of us have been embarrassed by some of the foolish messages that we have sent on the spur of the moment. It took a while for people to learn that some messages are best sent after taking time for reflection, some are best delivered in person, and some are best not sent at all.

It was with some trepidation late in 1979 that I agreed to the placement of a terminal on my desk, and I literally went through several hours of hand-holding with the Director of Data Processing before I ever logged on to the VAX by myself. After a few trials, I began to realize the power of the machine for my daily writing activity - correspondence, memoranda, reviews and comments, and - most important - manuscripts. I soon realized, also, that it no longer made any sense to prepare the documentation and proposals for my large-scale, longitudinal survey research project on paper; all of that material could be drafted or entered "on-line" and saved, revised, or reprinted as needed without any of the bother of photocopying and "cutting and pasting" that had taken so much time over the years.

If the isolation and privacy of distributed, interactive computing was an obstacle to the creation of an initial cadre of literate programmers and graduate students, it was an advantage of sorts in the diffusion of competence among the Center faculty. Many of us were incompetent at computing, and our fear of the machines was to a large degree a fear of appearing foolish when we made one mistake after another or just could not get something to work. The availability of a terminal on our desks, on-line HELP facilities, and - as a last resort - written documentation, spared us public embarrassment as we learned to use the VAX. By no means is this to say that everyone in the Center has become proficient in statistical computing or any other particular function of the machine; rather, we have been free to experiment with whatever functions of the VAX might appear to be useful in our personal style of work.

For example, once I became comfortable with the functions of the terminal as a writing tool, I became curious about its usefulness in statistical analysis, and I found it relatively easy to learn to do many of the things that I had previously delegated to Research Assistants - but with fewer communication problems and much faster "turn-around" time. I did not (and still have not) attempted to do any "data processing" as such. Rather, I have continued to assign Research Assistants the tasks of file management, file creation, and the preparation of cross-tabulations and correlation or covariance matrices. For myself, I have reserved the fun of multivariate analysis using programs like LISREL or GLIM.

One disadvantage of this change in my work patterns has been its effect on my training of graduate students. In the past, my complete dependence on student assistance made it necessary for me to work with students at every stage of data analysis. Obversely, the students observed every stage of the analysis and were able to participate up to the limits of their training and ability. This is no longer the case, and I have not yet figured out how to bring research apprentices back into my work without either boring myself silly or depriving myself of the real satisfaction of doing my own analysis.

**Data Management Utilities and Statistical Software**

At the core of our repertory of computer programs is a collection of utilities, many developed locally, for importing and exporting data, for managing large and complex data files, and for merging, extracting and recoding subsets of cases and variables from large files. The development of these utilities has made the data and statistical resources of the Center more accessible to users without skills, they have reduced the need to hire programmers for specific research projects, and they have increased the efficiency of our use of the computer. For example, we have found that none of the standard statistical packages is efficient in preparing data extracts from very large files, and we have developed dictionary-based extraction programs (XTRACT and HXTRACT for rectangular and hierarchical files) that are many times faster than SPSS. Similarly, we have developed our own "table-maker," XTAB, whose output is not so pretty as that of the statistical packages, but has vast gains in efficiency when we work with extremely large files, like those from the Census Public Use Microdata Files. We have developed our own set of symbols for tape processes that replace cumbersome VMS commands - as one of our staff puts it, DEC has yet to learn that tapes are used for anything other than system backup. We have developed an elegant little package of data management utilities (ODH for Our Data Handler), which simplifies such tasks as reading foreign tapes, reformatting (flat) files, selecting cases and variables, and producing frequency and character overviews. In addition to these programs, which were developed by staff and are fully supported, we have a collection of utility programs that have been contributed by users and are fully documented - usually with on-line help - but are not supported by staff. These include, for example, a local variant of Clifford Clogg's CANOAS program, the host portion of KERMIT, and utilities for marginal adjustment of contingency tables, random selection of cases within data files, and collapsing dimensions or categories of contingency tables.

Four or five years ago, we acquired the commercial version of INGRES, which is one of the major relational database management programs. Several projects have used INGRES, for example, to inspect the editing of records from 1/100 samples of the 1940 and 1950 U.S. Censuses of Population that have been selected and recoded under the direction of a team of CDE members; to manage and extract data from continuous life-history samples; and to archive, document, and cross-reference the documentation for a 20-year longitudinal study of Wisconsin youth; to change units of analysis in complex survey or census files; to match records across surveys; to manage the holdings of our data library. A major current research project within the Center is the SIPP Data Access and User Network, which has used INGRES to organize a user-friendly application for storing and accessing data from the Survey of Income and Program Participation. We are currently considering the possibility of a research development project that will combine relational data storage with optical storage media for a wider range of our data acquisitions and holdings.

In varying degree, the Center supports several commercial programs for demographic and statistical analysis. These include SPSS-X (Statistical Package for the Social Sciences) - which has been the real workhorse for some years - and MINITAB, GLIM, and BMDP, roughly in order of decreasing usage. We have benchmarked SAS from time to time and have always found it too slow for our liking. In addition to these multi-purpose statistical programs, we support several more specialized statistical programs: CENSPAC for the management and tabulation of Census files; LISREL, EQS, and HOTZTRAN for maximum likelihood estimation of structural equation models; ECTA, GLIM, MLOGIT, BIPROBIT, LIMDEP, and RATE for the estimation of loglinear and related models and for the estimation of continuous-time, discrete-state stochastic models.

The latter sets of programs exemplify a shift throughout the social sciences to more complex statistical models that cannot be estimated directly - as in the case of standard linear models like regression or analysis of variance - but require iterative estimation using numerical methods. This is one of the reasons that vast increases in computing power have not saturated our demand for more of it.

**Interaction, Batches, and Queues**

It was rather easy to ration access to the CDE computer before we got the VAX, for it could perform only one task at a time. Users simply signed up for blocks of time on a first-come, first-served basis, and the Center staff occasionally intervened to keep someone from "hogging" the machine. Access became much more complicated when the VAX arrived. Many users could work on the machine simultaneously, but would place different types of demands on the Central Processing Unit (CPU) and on input and output devices (disk drives, tape drives, terminals, and printer), depending on the tasks each was performing. The computer itself required "fine tuning" to handle this variety of tasks efficiently.

In interactive computing, the user initiates the job, say, a series of statistical computations, and no other use can be made of the user's terminal until the computer is finished responding to that command. In batch mode, the user instructs the computer to enter a job in a queue, and the user can do other work while waiting for the job to be executed and the output to be written to a disk file, tape, or paper. Usually, in either mode, a user directs output to a disk file, because the editor can be used to inspect it, and a decision can be made about whether to delete, hold, or print the file. This saves a lot of paper. If output is directed only to the terminal screen, the user gets just one chance to look at it without rerunning the job.

Some jobs obviously must be run in batch mode. Thus, we immediately created queues for printing, one-drive tape jobs, and two-drive tape jobs. Almost any other "job" except the entry of data directly from a terminal keyboard could be run either interactively or in batch mode. The computer gave priority in its assignment of CPU activity to jobs that were run interactively, rather than in the batch mode. Consequently, if a user initiated a "CPU-intensive" job interactively, it would monopolize most of the CPU cycles, and performance of the machine would decline for other users. This was readily visible to anyone using the system in the form of a longer wait for the machine to respond to any given command. So long as any user was free to begin a large, long, CPU-intensive job at any time, anyone could tie up most of the system and degrade performance for other users.

In order to solve this problem, we gradually developed a system of "sanctions" that encouraged fair and efficient use of the machine from batch queues while continuing to take advantage of the interactive technology for work that could best be initiated from a terminal. Currently, we put a limit on the share of CPU cycles used by an interactive process; when use exceeds 15 percent of available time during peak periods of use, a mail warning is sent to the user. However, this still encourages the use of the interactive mode for editing files and executing short programs. Thus, the entry of text, data, or programs; the testing of larger programs; and the execution of small-scale statistical analyses are still possible in the interactive mode. Second, we have developed a system of batch queues for larger jobs, and of priorities for running them, that vary with the length of the job (as specified by the user), time of day, and day of the week. Basically, short jobs are given high priority during prime weekday hours, while longer jobs - including long tape and print jobs - are given high priority at night and on the weekends. For example, during "prime" hours, we have non-tape queues with a maximum CPU-limit of 1 minute, 5 minutes, and 15 minutes; longer jobs can be submitted, but will not be run until 18.00. Recently, to encourage users not to hold data files on disk for too long, we have created a "SHORTTAPE" queue with a 5-minute CPU limit. In the evening and on weekends, a job can run for up to 4 CPU-hours without special arrangement with the staff. Most jobs can be "turned around" several times in the course of a day or evening of work, and we are usually successful in guaranteeing that any job will be completed within 24 hours.

Two other problems of rationing access deserve brief mention: disk space and efficiency. First, many jobs require the temporary use of very large amounts of disk space, and projects occasionally compete for these. We have experimented with several arrangements to deal with this; one was the allocation of a very large segment of disk on which anyone could write, but which would automatically be erased at 7.00 every morning. Some large users found ways to subvert this mechanism, while others found fault with the timing of the daily "purge" of the scratch-disk. We now permit any user to temporarily increase his or her personal disk space allocation, subject to severe and automatic sanctions if the space is filled for too long a period of time. The basic sanction is a delayed log-in time, the amount of time it takes for the system to respond when you turn on the terminal. Moreover, information about excessive disk usage is automatically sent to the leader of each project.

Second, there are many ways to accomplish the same task on a computer, and some are more efficient than others in the use of system resources. For example, a very large data extract can be prepared easily by a novice user of SPSS, but it may take 10 times more CPU-time to do the job that way than using our XTRACT program. Similarly, the system resources needed for an iterative statistical procedure may vary drastically, depending upon the skill of the user in choosing a set of "initial values" from which the computer searches for an optimal solution. Inefficient use of the system may be optimal for a user, but it is not for other members of the Center. We have no specific guidelines for dealing with this problem. Rather, the staff of the Computing unit (and some members of the Steering Committee) attempt to monitor the pattern of use by projects and individuals. The staff offers consultation to users who know they have problems - but those are usually not the worst offenders; it offers occasional short courses in various aspects of machine use. For the past two terms, we have offered a semester-long introduction to social scientific computing, which appears to have been very successful in introducing new users to the system. To supplement the blackboard and paper lectures of earlier years, we have equipped one classroom with a VAX port, and we have put a PC clone equipped with a Kodak DataShow on a cart. This has made it easy to demonstrate VAX features, terminal operation, and PC-based programs - as well as to disabuse novice users of the mistaken impression that mistakes are fatal or uncommon.

Our problems of rationing computer access are complicated by the fact that we have no price system as such. "Cost shops" charge for each operation on the system, and priority is governed by ability to pay. The major cost of inefficient use is borne by the user. In our system, we sometime vary the priority of certain classes of jobs, but all users are created equal. Control is purely normative; the only negative sanctions are log-in delay, peer pressure, and - ultimately - denial of permission to use the system. Operation of the system on this basis has required constant attention by the computing staff and the faculty, and its costs in our time and energy have not been small.

**Growth and Funding**

Over the past 8 years, we have modified the VAX and its peripherals in many ways in response to our changing needs and our experience with the machine. One of the first rules of computing is that the availability of a machine creates more demand for it! There are now 8 megabytes of main memory, 80 ports (which may support 45 interactive processes at once during a typical work-day), and 2.7 gigabytes of hard disk storage. We have replaced the original two tape-drives with three tri-density Systems Industries drives. Despite our efforts to upgrade this machine, its use reached the saturation point three years ago. We have made several efforts to supplement it, and we hope to replace it within the next two or three years.

When we purchased the CDE VAX, there was only one similar machine "on" the Madison Campus, and it was actually in our Physical Sciences Laboratory, an installation some 20 miles away. We rented a small amount of time on this machine while we waited for our VAX to be delivered, and we were thus able to adapt some of our software before the fact. After our VAX was installed, there was a tremendous rush on the part of people all over the Madison campus to redefine their work as "demography." We continued to spend a great deal of time - through a subcommittee of the CDE Steering Committee - in evaluating applications for Center membership. It was often difficult for us to draw the line between defensible and indefensible claims to be doing population research, and this difficulty was all the greater because of our interests in expanding the boundaries of population research and of helping our colleagues get their work done. On the other hand, we have had to weigh the expansion of Center membership against the legitimate policy concerns of NICHD and the evidently limited capacity of the CDE VAX.

During 1982, MACC purchased its own VAX 11/780, which it used to experiment with a new policy for selling computer services on campus. Rather than pricing services on a functional basis, MACC simply sold ports at a fixed annual rate, each supplied with a fixed allotment of hard disk storage. That is, the user obtained unlimited computing through the rental port. With the hope of improving the services available to non-demographers in the Sociology Department, we encouraged other members of the Department to pool their computing budgets and purchase 3 ports on the MACC VAX; these were accessible only by telephone. Although MACC was pleased with this operation and subsequently acquired a second VAX to be run in the same fashion, it proved a mixed blessing to members of the Sociology Department. On the positive side, it provided access to a very powerful machine at a fixed cost. On the negative side, the management of the MACC VAX was far less sensitive to the problems of social scientists than, by obvious contrast, the management of the CDE VAX. There was simply not as rich an array of software or of support services; there was a lot of raw computing power, but little else. Moreover, it was difficult to schedule the work of a large community of student and faculty users through just three ports. This experience has provided further confirmation that the success of CDE was not been just in acquiring some fancy hardware, but in creating an organization that could use it effectively and fairly.

Fortunately, late in 1983 we found a way to extend our user community to the Sociology Department as a whole (without jeopardizing NICHD support of the Center for Demography and Ecology). Indeed, we were able to increase the computing power available to sociologists in CDE at low marginal cost. At the completion of an NSF-supported research project conducted by several members of the Center, we have obtained a second VAX 11/780. This machine was similar in size and equipment to the CDE VAX as it was initially equipped; it has more memory, but low density tape drives and no public telephone access. We obtained support from the University for remodeling and installation, an operating budget equivalent to the University's present support of non-demographic computing in Sociology - split among the College of Letters and Sciences, the College of Agriculture and Life Sciences, and the Graduate School - and the expectation that Department members would continue to bring in external support to enhance the system. We installed the two VAXes side-by-side, with a networking arrangement that permitted the exchange of most information between the two computers, while reserving certain functions for use only by the CDE. The Sociology machine is controlled by the Executive Committee of the Department of Sociology, but it is operated at incremental cost by the CDE. The marginal costs of licensing, maintenance, and operation for the second machine have been low, excepting extended maintenance of some obsolete components; the greatest costs were those of training and consulting with a new and inexperienced group of users.

We grew again, about two years ago, when one member of CDE became Director of the University's Institute for Research on Aging and purchased a MicroVAX II, installed on our floor, for use in communication and data analysis by his interdisciplinary research community. In this case, unfortunately, we were not wise enough to provide for additional staff support. Although this machine, like the Sociology VAX, has generated more additional computing power for Center members than immediate demands for its use, its presence has taxed the capacity of staff to maintain parallel systems.

**Central Staff and their Functions**

The computing staff of the CDE has grown over the years, but we have tried to minimize specialization of function as much as possible, consistent with reasonably efficient operation of the system. There are a number of reasons for us to discourage too much specialization in the staff. First, we believe that excessive specialization is inimical to our interest in maintaining people's vision of the enterprise as a whole. Second, the University does not pay well, and there is a good deal of turnover in the lower staff positions. Fortunately, we have experienced very little turnover in several of the key positions, and we have excellent people in them. Third, our organization remains relatively small and understaffed at the core. Typically, the core staff are only partly supported by the core grant, and applications programmers have often had project-specific as well as Center-wide responsibilities.

The CDE Data Processing staff are presently organized as follows. The Director is responsible for the operation of the system and for supervision of other computing staff. He develops policy in consultation with the Center Director and a two-person Computer Policy Committee. This position requires managerial and teaching skills, broad expertise in statistical computing, and real understanding of the scientific agenda of Center researchers. We have had three Data Processing Directors since 1978, when the position was created, and we have found that our needs have better been met by persons with substantial social-scientific experience than by a computer scientist. The effectiveness of the Director depends on his ability to serve as a high-level diagnostician and consultant in contact with a great many of the projects underway in the Center and on his ability to translate what he learns from these contacts into computing policy.

The operational staff of the Center are an operations manager, three staff assistants for operations - one of whom manages operators, one of whom manages wiring and terminals, and the last of whom manages user accounts - a systems programmer, a data librarian, a statistical consultant, and one or more applications programmers. In fact, for the past year, we have gone without a staff applications programmer, other than those working on specific projects, and the restoration of one or more such positions is a high priority item for us.

The operations manager is responsible through staff for the design, acquisition, and maintenance of hardware, for acquisition and maintenance of software, for the hiring and supervision of 7 to 8 student-hourly operators, and for management of user accounts. Operators are available to mount and dismount tapes and run print jobs 16 to 18 hours per day, but the machine is kept running even when no operator is present. The systems programmer maintains, documents, and updates the operating system and other software and monitors use of the system on a daily basis. The data librarian - whose training is in library science - is responsible for the acquisition, documentation, and maintenance of a data library that now numbers about 6,000 reels of tape. As a practical matter, for many users she is the main point of contact with the computing staff and the main source of wisdom about getting things done. The statistical consultant is available to teach and work with users of a variety of "canned" programs and has some responsibility for finding, evaluating, testing, and adapting new software. Except in some cases where applications programmers have been hired by specific projects, they work on the development of new utilities for the CDE as a whole. That is, it is more common for project programmers to develop system-wide utilities than for staff programmers to do custom work for projects.

**Telephone Access**

Most ports on the CDE VAX are wired directly to faculty and research offices in our building, but several are connected to 1200/2400 baud modems and telephone lines. Although phone ports were initially installed to assist remote users in the Department of Rural Sociology, which operates an Applied Population Laboratory, or in the Wisconsin Survey Research Laboratory, Madison's "in-house" social survey organization, their primary purpose is now to permit remote, home access by faculty, staff, and graduate students. In addition to the public phone ports, many of the faculty have installed modems and A-B switches in their offices, which permit their phone lines to become dedicated ports in off hours. Such facilities are complemented by University tolerance of faculty who duplicate terminals or PCs for home use.

**Editing and Word Processing**

We were ill-prepared for the effect of the word processing capabilities of the VAX. We had only a standard line-printer on the system when the VAX was installed, and it printed only upper-case letters. One well-supported project purchased a "letter-quality" printer some months later, and its operator was soon swamped with requests to produce high quality documents for many users of the system. Now there are a large number of printers of varying quality and speed scattered throughout the Center: 2 high speed line-printers in the main machine room, a high-volume laser printer - equipped with LATEX for technical formatting, smaller laser printers in the administrative offices of the Center and of the Sociology Department, and in some faculty offices, letter-quality printers in project offices, and lower quality dot-matrix printers attached to the terminals of individual faculty members. In my opinion, we erred in deciding to install a central laser printer, rather than several, lower-priced and more widely distributed lasers; and we erred again in making LATEX the default formatter. LATEX is a wonderfully powerful and flexible system, but it is too cumbersome and time-consuming for applications that do not require its technical features.

Throughout the history of the Center, most faculty and staff have not had personal secretaries, but have shared the time of a few staff secretaries. This was not out of preference, but because we did not have the money to do more. The introduction of word processing made the use of a secretarial pool a far more attractive arrangement; in fact, I doubt that anyone in the Center could now generate enough work to occupy a secretary more than half time. The other side of this story is that few of us now have a secretary who really knows who we are and what we do.

**Problems and Prospects**

The continuing success of our research co-op is due in no small measure to practices of resource pooling and of normative social control; these have been built on trust and nurtured through daily personal contact and collaboration. There is a serious question whether these forms of resource allocation and social control can be maintained under the continuing influence of widely distributed computing power. The privacy and portability of interactive computing make it all too easy for faculty to work at home and appear at the University only for classes and office hours; to shut the office door and avoid wandering through the hallways; to find graduate students of little use or interest in the conduct of research. Each of these possible effects of distributed computing entails some change in the pattern of social interaction that has prevailed at Madison, and each is destructive to some of the sources of our success. Whatever the resources of its computing environment, for an academic research organization to be successful in the long run, its members must continue to find reasons to talk with one another.

But we have more concrete problems as well. In a nutshell, these all revolve around questions of replacement, acquisition, and allocation of computing resources. Our two 11/780s are each about 8 years old, overloaded during most hours of the day and night, and increasingly difficult to maintain. Although we need more CPU cycles and data-handling capability, our staff is overtaxed, and we are not likely to obtain many new positions. One obvious response to these problems is to move to distributed computing - remote PCs or workstations. We have tried this to some degree, though we joined the PC stampede rather late in the game. We were so well equipped and networked through our minicomputers early on that there was no great rush for us to buy PCs; little more than 3 years ago, there were only two PCs in the co-op. Now there must be 50 or 60 PCs around the place, and most dying terminals are being replaced by 8088 or 80286 machines. We are now acquiring our first VAX workstation 2000, initially to permit experimentation with local area clustering software, and no doubt we will find more workstations on the premises before the year is out, not all of which will operate under VMS. But the fact remains that, for our purposes, central computing remains essential. We are basically a data management operation, specialized in the analysis of large and complex data files, and data serving is likely to remain a "central place function" in the future computing environment.

Within the next few months, we plan to purchase two more MicroVAX IIs, at least one of which will be equipped with a 500 megabyte disk drive. These should relieve the present squeeze on interactive computing, be relatively easy to manage as members of the local cluster, and permit us to experiment with alternative mixes of interactive and/or specialized batch computing.

One of our major problems, once these machines are in place, will be structure a pattern of access and use that will increase throughput at minimum cost in additional software licensing and maximum transparency to users. There are a variety of uses or mixtures of uses of these machines with which we will want to experiment, and it is not yet clear to us what the best plan will be. One possibility will be to move batch jobs of long duration, but with minimum need for tape input, to one machine and to put heavy interactive users on the other. This would provide improvement in interactive response and batch response during daytime hours without requiring access across the network to tape drives. Another alternative may be to put as much interactive use as possible on each of the new machines during the day and to reserve the CDE and SOC machines for batch jobs, whether or not they require tape access. In either event, it is likely that we will use both of the new machines for long, CPU-intensive jobs at night.

As part of the competing renewal of our core Population Center grant, we plan to phase in a replacement for the CDE 11/780, providing a higher capacity data-serving machine, while combining the best parts of our old 780s in a "new" secondary data-server. We think this will help us to get the most out of the 780s in the remainder of their useful life.

Finally, we want to anticipate the replacement of tape by optical storage media, and of the mixture of terminals, PCs and VAXes with a combination of data-servers and powerful, desktop graphic workstations. As yet we see no emergent standards, either in optical storage and retrieval mechanisms, relational data storage and retrieval, or in desktop workstations for the social sciences. For example, will the latter be VAXes, PS/2 model 80s, MAC IIs, or UNIX-based workstations? Again, we hope to explore these possibilities within a development component of our NICHD renewal grant; we anticipate this will lead to a major investment in new data management and analysis technology within three to four years. Ultimately, we expect to build a computing environment where data will be transferred from optical disk to one of the data-serving machines; initial extraction, record construction, and recoding will take place on that machine; and analytic files will be transferred, transparently to the user's desktop workstation for further management, display, and analysis.