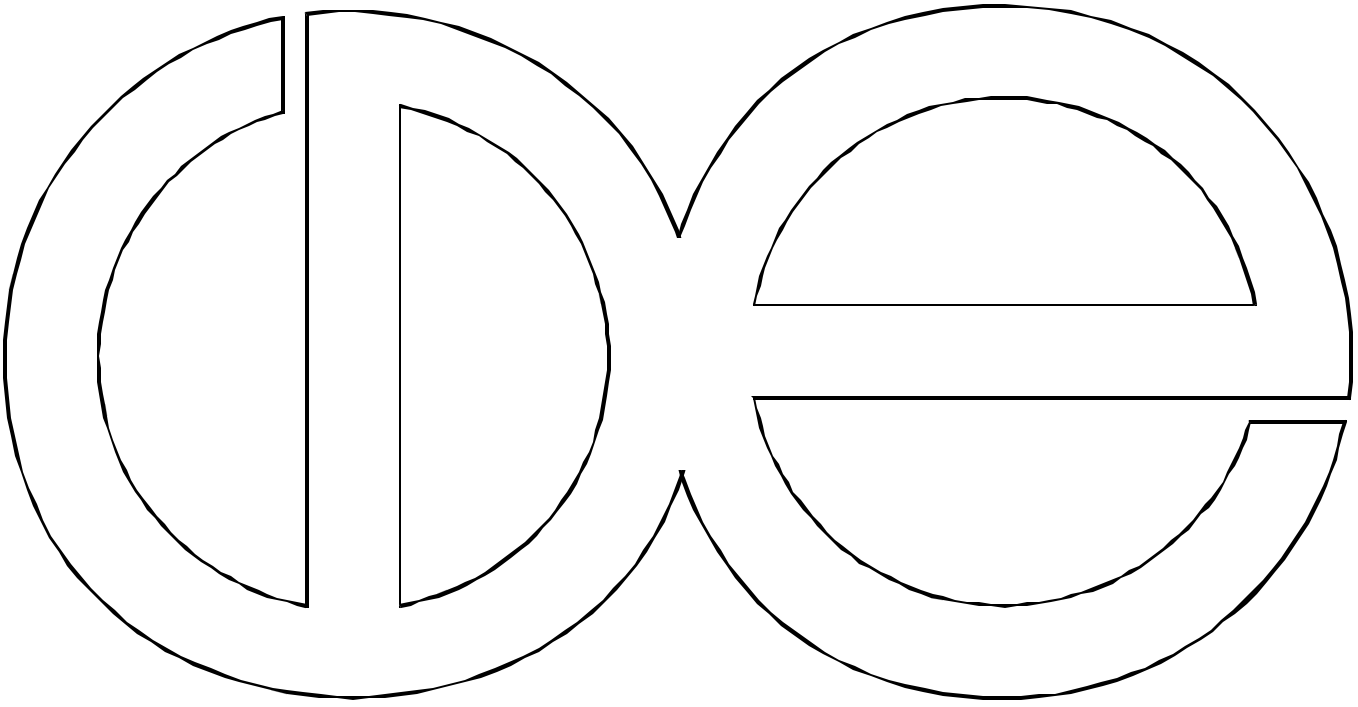
Center for Demography and Ecology

University of Wisconsin-Madison

A Cooperative Mode of Organization for Social Science Computing

Thomas S. Flory
Nancy J. McDermott

CDE Working Paper No. 99-30



A Cooperative Mode of Organization for Social Science Computing

Thomas S. Flory
Social Science Computing Coop
University of Wisconsin
Madison, WI 53706
(608)262-0499
flory@ssc.wisc.edu

Nancy J. McDermott
Social Science Computing Coop
University of Wisconsin
Madison, WI 52706
(608)262-3206
mcdermot@ssc.wisc.edu

ABSTRACT

Decentralized computing units have proliferated at many universities where there is also a presence of a large centralized computing unit. These smaller local units emerge for many reasons not the least important of which is in response to a perception that local computing and user support needs were not being adequately met by the central computing unit.

This paper examines a unique organization of a local computing unit at the University of Wisconsin-Madison. This unit, called the Social Science Computing Cooperative (SSCC), provides computing resources and user services for several social science departments and research institutes on campus. What makes the SSCC unique is the way it is funded. The SSCC does not get lump-sum payments from its sponsors that it can spend out of its own accounts. It gets no direct financial support from the university or the college. Nor does it operate a charge-back system for recovering costs as they occur.

This paper also looks at the relationship between the SSCC and the campus's central computing unit called Division of Information Technology (DoIT). In the provision of services like training, documentation, and help desk, which DoIT offers, it has proven beneficial for the SSCC to provide for itself to a great degree. On the other hand, in the provision of other services like software licensing, PC repair, and dial in modem pools, SSCC relies on DoIT.

Keywords

Decentralization, funding, social science research.

1. INTRODUCTION

A decentralized computing facility may or may not be the result of decentralization, intentional and managed. It could also be the result of neglect, separatist urges, irresistible economic realities, departmental entrepreneurship, or desperation. Another way of putting the matter is that the origins of such a facility may be in policy administered from the center, whether or not a department likes it, or in needs perceived by a department, whether or not the center likes it.

Needs differ, origins differ, organizational structures therefore differ, so how can we generalize about common challenges? To begin, we can get to know each other. In this paper we describe the peculiarities of the origins and operations of the Social Science Computing Cooperative of the University of Wisconsin-Madison. The SSCC developed out of needs that were felt by a particular research center and were not being met by the university's central computing facility. There are elements of neglect, separatism, economy, and entrepreneurship in the story and the result at this stage in the SSCC's history, a quarter of a century after its founding, is a unique organizational structure. Or so we think. Let's compare notes.

2. BACKGROUND

In the late 1960's the university's central and only academic computing center changed its main supported platform twice. The inconvenience and expense of learning new operating systems and hiring programmers to revise code wore down patience and budgets enough. But, having settled on a Univac 1100 mainframe by about 1970, the central facility in due course also changed its chargeback algorithm to the great disadvantage of researchers, like social scientists, who typically store and pass large amounts of data.

In 1972 the Center for Demography and Ecology (CDE) had had enough of this and responded by renting its own IBM 360 model 30. It had little storage but an industrial-strength tape drive. CDE used the 360 mainly for data reduction. Summary statistics were punched onto cards and the model-fitting continued to be done on the central mainframe. For a decade or so, decentralization was

tactical, guided by the relative costs of the different components of automated data processing.

By 1980 minicomputers had begun to change the computing landscape. CDE replaced its rented IBM equipment by purchasing its first Digital VAX. This was an 11/780, the classic Corvette of VAXes. Of course it had a small fraction of the speed, memory, and storage capacity of today's low-end PCs, but it was revolutionary nevertheless. There was no need to resort to the campus mainframe; decentralization became strategic: it was a defensible end in itself. Software was lean, sampling was a way of life, serial terminal connections became ubiquitous, and the interactive, timesharing model of distributed computing released spontaneity.

Through the support of its NICHD center grant, CDE was able to make its VAX available to all its members for all their computing needs without charging per-job costs. Certain inhibitions fell away. There was no marginal cost either in money or effort to testing new hypotheses or analytical methods, nor would a runaway job ever again break the bank.

Over the last generation every five-year era seems to have had its own view of what is liberating, what is distributed, what is efficient, and what is pacesetting. Timesharing on an early VAX was all these things partly because of advances in computer engineering, even more because of advances in operating system software, and most of all because it was now feasible for a small community of scholars to supply its own computing needs. There was scarcely an Internet, but there was intradepartmental email and a terminal on every desk. Technical advances made data processing easier and cheaper, but decentralization gave a new vigor to "computerized" social science research.

3. MORE RECENTLY

As other social science departments and institutes observed CDE's success – that is, both its increased research productivity and its ability to manage its own computing operation – they decided to follow suit. In the second half of the '80s four other units joined with CDE in an informal consortium. These units were the departments of Sociology and Rural Sociology, the Institute on Aging, and the Institute for Research on Poverty. Three of them brought their own new VAXes into the arrangement.

At first the VAXes were networked in a loose way, but there was no sharing of peripherals between systems. In short order, though, the new center adopted Digital's VAXcluster technology, a resource-sharing system that is not even close to being matched in elegance over a decade later by clustering innovations in the UNIX and NT worlds. This was the beginning of what was soon to become one of the biggest VAXclusters of its kind in the country. More important, it was the technical expression of a limited cooperative spirit. Decentralization as an organizational idea had run its course.

Probably a majority of individuals who now had access to resources in the new center belonged to more than one of the five sponsoring units, some to three or four of them. This overlapping membership pattern along with blurring of ownership boundaries created by the possibility of a general sharing of all clustered systems and the obvious economies of scale made cooperation seem sensible. From that time to rather recently, however, there has been very little expansive spirit and rather little demand from non-member units to be let in. In 1996 a sixth member joined, the Social Systems Research Institute of the Department of Economics. Then in 1999, the Department of Political Science joined.

The economies of scale in this operation can be illustrated by comparing the expansion of systems and services to the unchanging headcount of center support staff. Starting with CDE's single VAX 11/780 the VMS infrastructure alone grew into a 28-node Ethernet-connected VAXcluster. At this stage a large majority of the VAX machines were individual workstations purchased with grant monies and located in faculty offices. This trend followed a major investment by the university in upgraded cabling plants in academic buildings. The VAXcluster at this scale was fragile and unwieldy but was supported by the same number of technical personnel as had worked in the CDE computing operation ten years earlier. There were fewer programmers and more system managers, but the center had also acquired perhaps ten UNIX servers and workstations by this time and had taken on responsibility for managing the building's new Ethernet network and provided light support for the personal computers that were popping up in alarming numbers on faculty desks.

4. COOPERATION INSTITUTIONALIZED

Cooperation soon had to be formalized for two main reasons. There had to be a consensus method of determining how the operating expenses of the computing consortium would be shared; and there had to be an orderly flow of authority from the sponsoring agencies to the computing staff. The informal consortium of several years vintage became the Social Science Computing Cooperative when it was chartered by the College of Letters and Science in 1990.

By this means the SSCC got an official identity but it did not get and still does not have any fiscal autonomy from its sponsoring agencies. That is, the SSCC does not get lump-sum payments from its sponsors that it can spend out of its own accounts. It gets no direct financial support from the university or the college. Nor does it operate a chargeback system for recovering costs as they occur. Instead, at the beginning of each fiscal year it gets a commitment from each agency that it will spend a certain amount over the year on behalf of the SSCC. Each agency has identified money it can use for this purpose but in each case it may come

from different sources and be spent on different things. This is another sense in which limited cooperation has realized economies of scale. None of the agencies has money that has been allocated to it for support of the SSCC, but one may have money it can contribute toward staff salaries and another may be able to buy supplies, and so on.

The Chairs and Directors of the now six sponsoring agencies gather once a year, in the summer, to review the SSCC's proposed budget for the coming fiscal year and to decide amongst themselves which of them will take responsibility for which costs. The Chairs and Directors are presented with statistics on usage of SSCC resources by members of their agencies to help them decide on each agency's fair commitment for the year. But the expectation over the long run is that each agency will have lean and fat years, so that ability to pay is as important a part of this process as statistics on past usage. The cooperative spirit is intended to supply the patience that may needed when an agency learns that this year's contribution will have to be bigger than it expected. Somewhere down the road compensation will be made. The SSCC's history may not be long enough yet to evaluate this part of its operating philosophy, but so far it has worked. Three or four agencies have not yet had simultaneous lean years.

The SSCC staff are accountable to a second committee, the SSCC Steering Committee, made up of a faculty member from each of the six sponsoring agencies. Each member of this committee is appointed by his or her Chair or Director. The Steering Committee elects its own chairperson. The Chair of the Steering Committee sets the agenda of its monthly meetings and also acts as the supervisor of the one member of the SSCC staff who is designated as its director. Other staff report to the director. The Steering Committee considers issues of policy and represents the interests of the faculty and graduate student users. It mediates between users and staff in a very effective way. The nightmare possibility that the computing staff would find they each have six bosses has been altogether avoided, and in practice the work priorities of the staff are set most often by personal communications between the staff director and the chairperson of the Steering Committee.

5. ORGANIZATIONAL CHALLENGES

Underlying the formalities of organization are some challenges: one that is peculiar to this organization and one that must be shared by all "decentralized" computing centers. {scale}->{money's worth}

The SSCC is probably unusual in its cooperative organization. This sort of organization certainly does not imply, however, that the sponsoring agencies are not individually very concerned with fairness in the allocation of costs. Together, the coming year's proposed budget, previous years' levels of contribution, and statistics of usage aggregated for each agency set the terms of the discussion at the annual Chairs and Directors budgeting meeting. But the SSCC's budget doesn't change much from year to year and going into the meeting the participants are well aware already how much they have contributed in the past (and perhaps some of them have the feeling it was a bit too much). So the information that is examined most carefully as an indicator of how contributions might change is that on the patterns of usage of the SSCC's services and resources.

Usage accounting has been a deep and vexing subject since the beginning. At the very least an agency has the right to know what uses its own members are making of the SSCC in order to consider whether it is getting its money's worth quite apart from how much it is contributing relative to other agencies. The choice of measures is critical and difficult, but the value calculation is one that each agency has to make for itself according to its own lights.

The issue of fairness – the rightful size of each agency's contribution compared to that of every other agency – is touchier, not because agencies are exclusive and suspicious but because they are just the opposite. Since three of the agencies are multidisciplinary research institutes and two of the academic departments (Sociology and Rural Sociology) share a joint graduate program, a majority of all users of the SSCC's facilities belong to more than one sponsoring agency. A fair number belongs to as many as four. How are we to know how much of Professor X's computing should be attributed to CDE, the Sociology Department, or the Poverty Institute if he belongs to all three? And how do we measure his usage? What are the measurable indicators of a "big user"?

The measures of usage that have been tried are mostly those that VMS and UNIX accumulate by default about their own subsystems: things like I/O transfers, page faults, disk storage, and especially CPU time. In fact, CPU time alone is generally seen by the agencies' Chairs and Directors as the most intuitive measure of relative usage of the SSCC's facilities. It helps that this is a measure that the computing staff finds it easy to provide. However, all will admit that CPU time has in truth become almost meaningless in this context. At one time computing cycles were a scarce resource and most personnel were devoted to managing or developing software for the small number of available machines. By now the accumulated shared systems – mostly Digital Alpha machines running Digital UNIX – provide so large a store of computing power that CPU time is anything but a scarce resource. Likewise, disk space and memory are cheap and plentiful. Furthermore, the capital budget and the software budget, even the two together, make up only a small proportion of the SSCC's total budget. 80% or so of the budget, year in and year out, goes to staff salaries and benefits, and the majority of the staff are no longer involved in direct support of timesharing services.

Along with the problem of finding a way to measure usage, there is the issue of how to allocate the usage of those who belong to more than one agency. At one time we used an expensive

commercial package for what we called "project accounting". At login time users were asked what agency, or sometimes what research project, their upcoming work should be associated with. At the end of the computing session, the resources used during the session were partitioned into agency or project categories. But this system broke down mainly because most users couldn't be bothered to answer the original prompt accurately. One year we simply divided each user's usage statistics evenly among all the agencies the user belonged to. Nobody felt that was satisfactory.

Current practice:

- ask every multiagency user what percentage of his/her computing should be given to each agency.
- categorize every staff job into functional categories and allocate those to agencies where possible.
- total all the usage measures by user and allocate them to agencies according to survey results:
    a. CPU seconds of VMS and UNIX computing, weighted by power of machine
    b. Questions asked of consultant
    c. Attendance at SSCC courses

Problems with current practice:

- No CPU or other measurement of the use of supported microcomputers or PC servers.
- Hard to evaluate the use of some important user services such as documentation.

6. RELATIONSHIP TO REST OF CAMPUS

CDE anticipated a trend on the UW campus when it rented its first IBM 360 in 1972. In 1998 virtually all research computing is carried out in departmental centers. There are a few interdepartmental centers as well. That the SSCC is among the largest is an indication of how decentralized computing has become on campus. The university's central computing center is now called DoIT (Division of Information Technology) and is mainly in the service business, though it does still support two small VAXclusters for interactive timesharing.

Even in the provision of services like training, documentation, and software licensing, it has proved economical for the SSCC to provide for itself to a great degree. DoIT has arranged for some important campus-wide software site licenses, for SAS and SPSS for example. But the SSCC has negotiated its own site licenses for less widely used and less expensive software such as a nice telnet-based terminal emulation package and a PC X server, and it has collaborated with other centers on some licenses. DoIT runs a PC repair shop and sells PCs, Macs, and components; it maintains a very large dialin modem pool; and it maintains the campus backbone network. The SSCC performs complementary services in all these areas, including managing two small dialin pools and a high-bandwidth building network that includes ATM connectivity between Ethernet switches and the campus's own ATM network and full 10 MBPS or 100 MBPS Ethernet connections in every office.

Recognizing that one of its most important contributions to campus computing could be as a hub of communications between the many decentralized computing centers, DoIT organized an organization known as Tech Partners. Tech Partners exists primarily as a very active listserver by means of which the campus's scattered computing support personnel ask each other technical questions, discuss the challenges of a highly decentralized computing environment, organize participants in software site licenses, and so on.

The most interesting technical response at the university to the scattering of high-powered computers around campus is a CPU cycle-sharing software technology called Condor. Condor has been developed over several years by a group of computer scientists who saw in the decentralized geography of computing huge quantities of CPU cycles that were simply going unused. One researcher might indeed make good use of her 300 MHz processor from time to time. She needs a powerful computer, but she uses its full power only infrequently. The vast majority of the thousands of computers on the campus, any campus, are underused in this sense, but before Condor there was no way, lacking a central authority, to rationalize the distribution of computing power.

Without going into any technical detail at all, Condor simply pools the cycles in its so-called "flock" of participating CPUs. If one's PC isn't up to a particular task, Condor finds an idle machine that is and ships the job over the campus network to that machine. Indeed, it ships the entire computing environment to the remote machine. The remote computer does not have to have the same application software installed as the originating machine or even the same operating system. Most local networks outside the computer science and engineering departments have not been included in Condor flocks until very recently. Last year the University of Wisconsin was one of several American universities to receive grants of high-end computing equipment from the Intel Corp. in order to test the suitability of Windows NT on Intel-based machines for computationally intensive work. The SSCC was included in the university's award and will took delivery in the autumn of 1998 on ten client computers and one server. All of the several million dollars worth of equipment that the university was awarded will be included in the campus Condor flock.

Although the hundreds of existing machines running Condor already attest to the software's scalability, this will be the first time that Condor has been applied campus-wide. The main application envisioned for the SSCC's equipment is Geographical Information Systems but we also intend to run the usual social science statistical packages and we are extremely interested to learn whether cycle sharing at this scale will make it possible for all of the university's decentralized computing centers to save money on equipment purchases by cooperating at this purely

technical level.

On campus there is no obvious trend toward administrative recentralization of computing. The advantages of self-determination are too great for that to happen. The SSCC itself found some economies in bouncing back from decentralization to limited cooperation among like-minded social science units, but for now an equilibrium has been reached. Despite years of concern over the quality of computing support that its departments get, the College of Letters and Science has never regarded the SSCC, its only successful multi-departmental computing facility, as the seed of a College-level computing center or even of a broader community of operations among all the social sciences. This is a little curious. The entrepreneurial departments and those with a little discretionary funding can make their own arrangements, of which the SSCC is a rather unusual one. But without College support, departments that cannot pay for such tailored computing services are on their own.

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: flory@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu