Moving To Distributed Computing: Experiences From The Minicomputer Transition

by Halliman H. Winsborough [1]
Social Sciences Computing Cooperative
University of Wisconsin, Madison

Where We Are Going

In these remarks, I take the phrase "distributed computing" to indicate the expected computing environment of the next several years rather than its more technical and narrow meaning. Most of us will want to move in the direction of this expected environment in order to do our work with competitive efficiency. I think this environment has four elements:

1. Powerful processing is accessible from all users'desks. That power is likely to be many times greater than that available in the past. Many computers may be involved in making that happen. They are accessible from every desk that needs access. They are also accessible from home, hotel room, laptop, and - God help us - from the car.

2. Powerful connections are available from the user's desk. The desk top, lap top, car top, machine, whatever it may be, is connected through the electronic network to other machines locally and to the national and international electronic networks. Electronic mail is the "normal" mode of communication locally, nationally, and internationally. In principle, data at remote locations can be accessed easily. Software that is legally available to the user can be accessed remotely. The user may run programs on her own machine or on the remote machine. In the best of these ideal worlds, running locally doesn't requires recompiling.

3. Maintenance of all these wonders is invisible to the user. Machines are connected, repaired, and replaced. Files are backed up. Important new files are added and potential users informed of their availability. Documentation is maintained and improved. Programs are checked for accuracy. Network addresses are updated. Network protocols and even physical connections are changed. All this behind the scenes, as it were.

4. Openness prevails. There is standardization of operating systems, editors, and programs. As aresult, a user can work on a new machine or on a remote machine with only modest additional training. In the best of these worlds, standardization pertains to data as well as systems. In this world, one would retrieve data from, for example, Dialog, Cendata, and ICPSR using the same "language."

No doubt some of this description seems hopelessly utopian, even to the most enthusiastic among us. But a good deal of it is currently in place.

Powerful machines are here. We just proposed a Sparcstation 10 for a faculty member. At about $10,000, it will compute at 85 or so MIPS. That is mainframe speed. Several competitors do as well. But even that kind of power is not sufficient for one of the faculty members I serve. He routinely ships jobs from his desk to a supercomputer in San Diego.

Communications improvements abound. Electronic mail is a commonplace. I suspect the organizers of this conference wonder how they could have done their job without it. I also suspect that remote access to data is an ongoing theme in this association. I will have more to say subsequently about what we must do in order to make data access fit the new computing environment.

When it comes to maintenance and support, things get more speculative. A lot goes on without the user knowing about it but sometimes the behind-the-scenes machinery creaks pretty loudly and an occasional flyer falls on the cast. That is because distributed processing can get pretty complicated. The tangle of things can get so dense that it is hard to see the bug before he bites.

Openness and standardization is in process but not very far along. As of today, trends are mixed about how well this user-demanded principle will stand up to corporate proprietary urges. A year ago all the big players were marketing openness. But recent events suggest a retrenchment. The ACE consortium looks moribund. SUN doesn't even make a C compiler for its new machines, so it is a bit harder to eschew SOLARIS for BSD than it was. And so on.

Over all, then, there is a lot of progress toward the ideal distributed computing environment but a lot of room for uncertainty as well. As I visit my colleagues at other universities, I sense quite a lot of uneasiness about the transition from whatever kind of computing they currently have to the new environment. My informal survey

suggests that how awesome, impractical and distant the norm of distributed processing seems depends a good deal on where you start from.

Where We Are Coming From

People are facing the transition to distributed computing from a number of different current environments. All of those environments have elements of the future in them - some more than others. In the following I will distinguish three types of startpoint environments - mainframe shops, personal computer shops, and minicomputer shops - and discuss how the transition to distributed computing looks from each vantage point.

The Mainframe Shop

By a mainframe shop, I mean a group that depends on a large, centralized, computing "utility." People from this environment are used to quite powerful machines and find nothing very exciting about a computer that turns out 85 MIPS. It is what they expect. They are also used to a pretty high level of invisible maintenance and technical support; so good, in fact, that it leads to change-resistant users, as we will see. The organization of access to data in such a shop can be superb. But often it is not.

Connectivity is less familiar to people from the mainframe environment. It is unusual for everyone in a mainframe shop to have a terminal on their desk. Batch processing remains a main mode of work. Although interactive computing is available from mainframes, it is pretty pallid stuff. You go to a terminal to create and submit a batch job. IBM has introduced PROFS recently to permit local communication, but it doesn't have the same presence as e-mail does when everyone has a connected machine on their desk.

Openness doesn't exist in mainframe shops. Enough of them use the same vender's equipment, though, that movement from one mainframe shop to another is fairly easy. Thus monopoly substitutes for openness and the only victim is price.

I think it is people from mainframe shops who react most violently to the prospects of a transition to distributed computing. Those who haven't begun the transition are most resistant. Those who have made serious strides toward distributed computing are the most ecumenical. Partly, I think, it is because the mainframe mavens have done such a good job of making things transparent. In so doing, the mainframe priesthood has shielded social science users from the grubbier aspects of computing by making them appear an esoteric mystery - so much so as to produce a kind of learned helplessness in the users. An important part of making the transition to distributed computing is to take some things into your own hands. That prospect can look remarkably dangerous, even sacrilegious, to oldline mainframe users. Once converted, well, it is like the old saw. Besides, the new environment is worlds better.

The Personal Computer Shop

PC shops, until recently at least, aren't really shops. The big thing about a personal computer is that it is personal. It's yours. It's on your desk. You take care of it, buy software for it, install the stuff, decide when to upgrade the operating system and do it yourself, back it up, defragment its little disk, change the battery for its clock and install new boards, interfaces and disks. The idea of doing it yourself isn't daunting to people from the PCworld. It is just a bore. Invisible maintenance can seem like a dream, especially when your disk crashes and you realize you forgot to back up last night.

People from the PC world are also pretty comfortable with interactive computing. They expect "standards." They also are often quite interested in more computing power, sometimes to a level of fixation that raises my Freudian eyebrows.

I think it is the connectivity of distributed computing that gives PC people the most trouble. It is all so un-personal. Connectivity and the consequent standards reduce the user's freedom to do anything on "their" machine that they wish. But connectivity is beginning to catch on even here. Witness the success of Compuserve.

The result of all this is that PC users are a lot more eager than mainframe people to make the transition to distributed computing. Most PC people are eager to have a powerful, networked UNIX box on their desk. They just insist that the desk be big enough to hold their PC, too.

The Minicomputer Shop

In a classic minicomputer shop, users have terminals on their desk that are connected to a rather modest computer. Such shops start off closest to distributed computing. One accesses computing cycles from the desk. Computing is interactive. Communications with one's own work group are quite facile. Wider area access to cycles, data, and software has been in place for some years. Maintenance is pretty invisible. If your mini run UNIX, many of the things listed under my openness rubric were there, too. If you run one of the proprietary operating systems, such as VMS, openness has been a lot slower in coming.

Minicomputer types generally feel that the transition to distributed computing is just a bit more of what they have been used to for a long time. The big attraction is the increased power and, for those stuck in proprietary operating systems, increased openness.

One of the reasons that people from minicomputer shops face the transition to distributed computing with a bit

more equanimity than people from mainframe shops or PC shops is that they have already made important parts of the transition. Because of this history of change - this slower transition - the experiences of one minicomputer shop may be of some use in thinking about making the transition to distributed computing in other places.

The Experience of One Minicomputer Shop.
The Social Sciences Computing Cooperative at the University of Wisconsin, Madison, where I work, has been operating a computing facility for social science research since 1972. For the first eight years, we operated in the "mainframe" model; complete with glass enclosed shrine, an IBM iron god, and batch processing.

In 1980, we made the minicomputer transition when we gota VAX 11/780. Before long, nearly every faculty office and most of the research rooms had terminals connected to the VAX. We didn't (and still don't) charge for resources used. The mail system was pretty good. Suddenly our clients had copious interactive computing and were connected in an instant communications network. That was the most dramatic subjective transition that we have made.

We started technical distributed processing in about 1988 when we began to distribute tasks among several VAXes that were previously independent network partners. Now we operate two client-server UNIX systems as well as a Local Area VAXcluster - the direct progeny of the VAX 11/780 - and a growing PATHWORKS network of PC's and MAC's. From the latter, it is easy to connect to any of the former networks.

There are four aspects of these transitions that were somewhat unexpected for us. I pass them along in the hope that they will be of some help to those of you are just beginning the transition.

1. It costs a lot to service fancy equipment in people'soffices.

2. Teaching becomes an increasingly important activity.

3. Rapidly dropping costs means that plans and policiesmust stay flexible and be reviewed regularly.

4. The social organization of computing becomes asimportant as its technical aspects.

In the remaining pages I will discuss each of these findings as we experienced then. Then I will discuss problems associated with the transition we haven't made, the transition to distributed, on-line data.

Equipment in Offices.
When we moved from the mainframe to the minicomputer, our operations people proposed the policy that our responsibility for equipment should go from the machine room to the wall plug and no further. The terminal on the user's desk was the user's problem. Our organization has always been a consumer's co-op, so that policy lasted about a week. Diagnosing, repairing, and replacing faulty terminals became a standard task for us. Initially, the co-op provided fairly simple terminals. As time went on, people wanted fancier machines and bought them from grant funds. We took care of those, too.

As PC's became more popular, many users bought one for home. Before long they wanted to use terminal emulation software and call in from their home PC. So we got in the modem business and even took over some maintenance of home PC's. The emulation software worked well, and some users decided they wanted PC's in their offices rather than terminals. Some place in there we should have reared back and passed a policy about what kind of equipment we would service and what we wouldn't. But we didn't. So we got into the business of repairing nearly any kind of PC computer, printer, or storage device and ensuring that it worked in a civilized way with the other computers in the system. It was foolish of us. A faculty member saved $75 by buying an unfamiliar laser printer and we spent $750 in time making the thing work properly on our networks.

The advent of workstations brought some order to our policies. We decided that the co-op had to agree to service a non-standard workstation before its purchase or the user was on his own. We have extended that policy to other equipment as well. Of course, that meant we had to decide on what was "standard" in the pc equipment business. That is taking some time, but we expect it will have good results for both users and co-op staff.

Teaching
Teaching rather sneaked up on us, too. Initially, we gave occasional lectures as introductions to our systems and to provide some training on software we had written. Of course we have always provided fairly extensive consulting. Since we are in a university, we get a fairly large batch of new users every year. Before long we were doing more extensive training of new users - training designed to reduce the burden of answering the same question over and over again in consulting. Then the people who teach statistics decided they wanted us to take over more of the training in how to use the statistical software. So that got added to our teaching portfolio.

With the addition of UNIX to our operating system mix, we are doing more short courses in the operating system and its editors.

A new addition to the list next school year will be instruction in SQL. We have taught about relational ideas and data normalization for several years but instruction in SQL will be a new addition.

One result is that over the years we have added personnel in the consulting and teaching part of the staff. User Services, as we call these functions, are about 1/3 of our staff activities. We did not expect it to grow to such a large fraction.

Of course it would have been possible for our organization to have avoided doing many of these things. But they represent real user needs. If we didn't satisfy them, they wouldn't just go away.

Things are Cheap

It is wonderful that the price of computing equipment has fallen so dramatically in the past decade. Keeping up with the changes can be a problem for a computing organization, however. Not only do the people in charge of buying things have to keep their information refreshed but also one must re-think policies on a regular basis to see if they were made contingent on a particular price environment. Take disk space, for example. We initially allocated new users 2000 blocks of disk space on the VAX. That was when a 75 Meg disk for the VAX costs $20,000. It became a kind of rule of thumb that lasted much too long into the dramatic decline in disk prices. We now try to regularly review policies to see if they are outmoded. New employees can be especiallyhelpful in detecting these residues of previous price regimes.

The Social Organization of Computing

The flexibility of technical computing arrangements has grown so dramatically in the past several years and the price of computing has gone down so dramatically that we currently believe that the greatest leverage in computing efficiency can be achieved by using the new flexibility to modify the social organization of computing. Three organizational modifications have been particularly useful to us. First, we have become a consumer's co-op - user owned as it were. Second, we deal with money in a special way. We don't charge for computing. Co-op members agree to contribute to co-op costs from their budgets. Third, we use the flexibility of modern computing to "fit" the unique work-group style of users. We don't have much pride of invention about these arrangements. Like many opportunities for organizational change, they rather happened to us and we tried to keep the ones that looked promising.

Initially we were the computing arm of the Center for Demography and Ecology. In the mid-1980's, several other organizations on campus came into some computing money and decided they wanted to join with CDE in providing services to their members. Since there was a very considerable membership overlap between CDE and these organizations, it made a lot of social and political as well as economic sense to try to achieve the expected aggregation economies. The growing flexibility of computing made this organizational arrangement possible.

That's when we formed the Co-op. In this new organization, each of the "sustaining" organizations has a more or less equal say in what goes on. Policy decisions and oversight are performed by a "steering committee" made up of representatives from each agency. The budget is decided annually by the chairs and directors. It has worked pretty well so far. The non-faculty computing director has the committee as Boss. When agencies' needs conflict, he can ask the committee to decide how to play fair rather than making it up himself.

As you can see from the foregoing, we deal with money and accountability in a special way. Agencies decide each year how much they should contribute to the expenses of the co-op. Agencies own some of the machines that we run and pay the attendant software, maintenance, and supply costs for those machines. Other machines are held in common. Each agency pays a share of the cost of those machines. We have used the flexibility of the various operating systems to keep the permissions straight in this arrangement. Users are authorized on machines belongingt o agencies they are members of and on common machines. The accounting system keeps pretty good track of who's doing what on all the machines and what agency is responsible for the time.

The notion of common machines is more flexible than one might initially suspect. Certainly servers are common machines. But we also retain some older and smaller VAXes as common machines because software is cheap on them. We have them loaded up with software that is used only occasionally by any one group but is cost-effective to license on a small machine for the whole co-op's use.

Finally, we use the flexibility made available to us by distributed processing to fit a work group's computing as closely as possible to its special needs and style. For example, most co-op members have been fairly happy with our system for using tapes. Operators are on duty about 18 hours a day and do the tapemounts. The Institute for Research on Poverty, however, has a group of programmers that do quite a lot of work with large files- CPS and the like. They very much like to mount their own tapes. So IRP has a tape drive on one of its machines in a room accessible to its programmers and they do their own mounting.

Data Access in a Distributed Environment

The last issue I want to address is the one of data access

in the distributed environment. I think this is an issue we all face. A crude way of putting it is, "What will we ever do without round tapes?" Some people seem quite far along. Jim Jacobs with the social science group at San Diego has a wonderful jukebox/menu-interface. It is the neatest thing I have seen. Al Anderson in the Demography group at Michigan has a plan for data to be delivered from a campus data utility over local FDDI to a RISC machine with an enormous main memory space for buffer. It is the most ambitious thing I have seen.

In the co-op we are moving fairly slowly to rid ourselves of round tapes. At the same time, we aren't buying replacements for the nearly worn out ones we have. After several years of thinking, visiting other installations, and trying things out, we have come to an important conclusion for our shop. It was really Tom Flory's insight. It looks like the big issue is the media you will use next; whether to go to WORMS, MO's, DAT's, or 3480's. But that's probably unanswerable without knowing how you are going to use the equipment. We think that the place to start is with the interface. What should the user's access look like? What kind of tools for extracting data should be available? Do you need to do complex joins as well as restriction and projection? How frequently? How is information about the data to be coordinated with the access process? How is one to implement solutions to these problems in a way that is reasonably open and standard? These questions and the others that arise in answering them are bedeviling us currently.

When the only media was round tape, the answers to these questions were fairly constrained because serial access is fairly constraining. We can now debate about the most amazing things: Is it more "standard" to preserve archival provenance and keep the data in the form we get it from the distributor or is it more "standard" to rearrange and decompose files to satisfy, say, third normal form? Should we use a commercial data base, say Ingres, to organize the data and make relational joins possible? Or can we get along with what you can do in SAS and SPSS?

We haven't come to any grand solutions to these problems. We lean toward normalizing the files and keeping them as ASCII files. For the moment, our solution to the media problem is to buy quite a number of SCSI drives. We will keep the most used data online, probably in compressed form, on these devices. Our interface decisions will be made assuming that whatever media eventually is favored, it will be possible to make the machine think it is just another directory.

It is an exciting time for all of us in the computing business right now. It is probably most exciting for those of us who deal in data. For the first time, there are the facilities out there at a reasonable price for us to serve our users much more effectively. If we can now just manage to do it in an open and standard way, all will be well.

1 Presented at the IASSIST 92 Conference held in Madison, Wisconsin, U.S.A. May 26 - 29, 1992.

The Center for Demography and Ecology receives core support for Population Research from the National Institute for Child Health and Human Development (P30 HD050876).