The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors

The QUARTERLY is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press Manual of Style or Kate L. Turabian's Manual for Writers. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". Journal of the American Society for Information Science 30(2):77-82, March 1979. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor: Walter Provesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University, Burnaby, B.C., V3A 1S6 CANADA. (604) 291-4349 E-Mail: USERDLJR@SFU.BITNET

Book reviews should be submitted in duplicate to the Book Review Editor: Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California 92713 USA. (714) 856-4978 E-Mail: DTSANG@ORION.CF.UCL.EEDU

Title: Newsletter - International Association for Social Science Information Service and Technology

ISSN - United States: 0739-1137 Copyright 1985 by IASSIST. All rights reserved.
Moving To Distributed Computing: Experiences From The Minicomputer Transition

by Hallman H. Winsborough
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 require 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 a result, 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 table. 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 re-trenchment. 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 vendor’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 PC world. 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 got 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's offices.
2. Teaching becomes an increasingly important activity.
3. Rapidly dropping costs means that plans and policies must stay flexible and be reviewed regularly.
4. The social organization of computing becomes as important as its technical aspects.

In the remaining pages I will discuss each of these findings as we experienced them. 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 sneakily 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 especially helpful 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 belonging to 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 tape mounts. 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.

I 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).
A User’s Perspective on Electronic Data Archival: The Importance of Standards
by Annette Jones Watters¹
and Carl E. Ferguson, Jr.
The University of Alabama

Revolution probably wins the prize for the most over-used characterization of rapidly changing non-violent events. However, few words better characterize the rapid rise of the microcomputer as the dominant technology of the 1980s. The device has revolutionized the workplace, bringing the power of electronic digital computing to the desktop. And, with speed and storage capacity increasing extremely rapidly, the microcomputer continues to transform every task associated with the acquisition, maintenance, and use of information. This paper offers a brief look at the impact of the microcomputer revolution on data distribution and archiving standards, then attempts to chart current trends and conditions in the rapidly changing technological landscape.

Historical Perspective
Digital document archival has traditionally been directed by considerations of space and convenience.² Space was a consideration because traditional library or reference facilities simply could not accommodate copies of all the historical information. Although magnetic tape offered relative high storage densities, even tape storage quickly became problematic. The space savings achieved by going from tape canisters stored in wire racks to hanging tape seals was quite significant. However, every unit eventually ran out of room—no one could ever buy enough tape cabinets.

Convenience. Seldom used historical data could not compete successfully with current information for shelf-space. As a result, data progenitors and librarians soon developed usage rules to help establish shelf-life and retention standards for data sets.

The Limits of Space and Accessibility
Limited space dictated that out-of-date items be compressed and/or relegated to less expensive (albeit less accessible) mediums. Numeric data was frequently transcribed from a character format (typically ASCII or EBCDIC) to a much more dense binary format. The resulting files were then written to the highest density magnetic tapes available. Standard tabulations based on these data were candidates for microfiche, 35mm microfilm, or paper microform products. Binary tapes and microform products do offer significant storage densities. However, retrieval has always been a tiresome process.

In all cases, accessibility and space savings were the primary considerations and the end-user was frequently the loser. The end-user usually played little or no direct role in the determining the method of compression or archival.³ Indeed, the end-user generally worked through an intermediary who selected the archival strategy. That strategy frequently was not based on the needs or retrieval skills of the end-user. A critical aspect of these archival strategies was the skills and tools available to the person archiving the data, not the skills and tools of the researcher. Archival responsibilities frequently fell to computer programmers who had no sense of the practical value of the data involved.

Machine Time
Prior to the advent of the microcomputer, most data analysts worked with paper products developed and maintained by a group of modern-day alchemists called the programmers. Working patiently, the analyst communicated the nature of the application to the wizard, who with cards in hand communicated with the machine. This was a most serious relationship, for usually there was only one machine in the organization—one computer to be used by all. Competition for its time and attention could be intense.

Trial and error was expensive. Research strategies requiring alternative methods of analysis were expensive. Researchers were allocated a limited amount of machine time and they learned patience. They conceptualized the table or statistical procedure to be run, gave it to the programmer, and waited. Two or three turn-arounds in the morning—maybe the same in the afternoon—meant that they did not spend too much time trying alternative methods or procedures.

With the development of the statistical packages, SPSS, SAS and others, the role of programmer as the analyst’s interpreter began to fade—though not yet disappear. The canned packages greatly facilitated the analysis of these data and they confronted the analyst with a new challenge. The intellectual cost to the analyst in time and commitment to develop programming skills in FORTRAN or some other higher level language was almost always viewed as excessive. However, the statistical packages were different. Using surprisingly few procedural commands the analyst could now read the data and

¹º 1992
Spring/Summer 1992
9
actually do the analysis. Most researchers immediately realized how much time could be saved by skipping the intermediate step of using a computer programmer. That time might now be used to explore alternative methods or forms of analysis.

Consequence
It was the limits imposed by space, time, and accessibility, that profoundly directed the data distribution and archival standards of the 60s and 70s. Programmers working for data progenitors prepared distribution tapes for programmers working for data analysts. And, programmers chose the archival standards and formats for the day that someone would want to look at old data sets. End-users rarely read tapes. End-users worked through programmers and it was the programmers who decided how they would communicate with one another.

The Evolution of Desktop Computing
While Apple and others were offering microcomputers in the late 1970s, the introduction of the IBM Personal Computer (PC) must be regarded as the beginning of the workplace revolution. IBM’s entry into the market gave the microcomputer credibility. It was no longer a toy or experimental device for hobbyists — it was made for work and from the first day began to recreate the office. One could say, “And the rest is history!” but there is too much to be learned from this transformation of the workplace to move on too quickly.

At first office workers were given a machine, and little else. Many quickly learned two new words — hardware and software. They learned that without software the hardware did not do very much!

Software
These microcomputers were fast and could remember things! And they did like numbers. However, they were business machines — they liked documents and numbers. The numbers they liked best were of the financial variety (spreadsheets) and the documents were correspondence. Lotus, Microsoft Word, and others quickly found their way into the market — and the world would never be the same.

Hardware
As more software and data applications became available, the 10MB hard disk quickly filled up. Although the earliest microcomputer chip — the 8086 — was fast, more complex applications quickly called for more speed. Today, the fastest machine uses an 80486 INTEL chip running at 50MHz, 8 MB of RAM, and is typically packaged with a 350MB hard disk. Such a machine will operate hundreds of times faster than the original 8086 and offers more total computing power than large mainframe computers of less than a decade ago.

Socialization
By the end of the decade the VLSI (very large scale integration) silicon chip, that thumb nail sized computer, could be found in every office and on almost every desk. It was no longer a curio down the hall but rather an extension of the worker. In the decade of the 1980s it was OK for men to type and for senior executives to get their hands dirty with data.

Scientists captured data via analog ports while specialized software, running in the background, conducted the analysis in real-time. Survey research introduced CAI. SPSS, SAS, and BMD for the PC were not far behind. Never before could the analyst get so close to so much data — manipulate it, manage it, analyze, and interpret it. Microcomputer based analysis and text processing (eventually to be called desktop publishing) skills were fast becoming an integral part of every data user’s personal skill set. Whether the analyst was a social scientist, music historian, paleontologist, or Greek mythologist, the power of the micro was sweeter than the songs of the Sirens.

User groups, first formed to provide aid and comfort to practitioners of the infant technology, disappeared as help became available from the officemate next door. Power users began talking to software and hardware developers, offering (frequently demanding) new features, more power, more speed. And, as the size of the market continued to grow, the software developers listened; their craft was now a multi-billion dollar business.

And so, what has become of the programmer? Who now sets the distribution standards? What has become of the limits of space and time?

Microcomputer hardware, application software, and enhanced user skills have dramatically altered the traditional role of the programmer data analyst. The installed base of MS DOS microcomputers is now measured in the hundreds of millions and the market potential for a good applications software package can quickly exceed a million dollars. Microcomputer applications software developers have attracted exceptionally bright and creative systems designers and programmers with training and interests in many functional fields. As a result, researchers now have computer based tools unimaginable less than a decade ago.

Standards
The Uses of Standards
Electronic data distribution and archival standards serve the user community in several way. Standards promote case of communication among users and between users and data providers;
equitable global access to data opportunities through improved documentation and communication environments; and the convergence of distribution and archival media.

Ease of communications between data users and data provides is critical to both analyst and provider. Frequent providers include governments (national, state and local), universities and other research organizations, and businesses. While each has a unique mission in our society, as data providers, they and their user community can benefit from improved communications — improvements through mutually agreed upon standards.

The user community, public and private social and physical science researchers and analysts, share in this responsibility.

All too frequently, a me versus them mentality sets in. If there were a common understanding of the technical standard for providing data and common understanding of what is reasonable for the end-user to bring to the table, the level of antagonism would be reduced. These standards of expectation do not now exist.

Improved, jointly developed standards, are a major step toward equitable global access. Global communications today is no more exotic then a hard-wire link to your local mainframe in an adjacent building. However, to be most useful, data providers and user worldwide must work closely together to insure not just interagency or national standards but rather international (universal) agreements on media and form.

Such standards must transcend multiple platforms and operating systems. Microsoft DOS machines must be able to easily communicate with UNIX (XENIX), MacIntosh, and others. Communication standards are desperately needed to allow word processing (desktop publishing) software to easily share a document and its complete formatting. Microsoft, with its rich text file (RTF) concept is offering the market one such standard for consideration. And, of course, the need for standards can be found for spreadsheet, database, CAD, and other systems.

What has changed is the role of the user. The size and sophistication of the user community both commands the attention of the developers and shares with them the responsibility to develop and adopt global standards in each of these functional areas. Poorly written documentation and a lack of common understanding on what documentation is supposed to cover disrupts international data distribution, even without intervening language barriers.

The Future of Standards
While it may seem contradictory, standards are dynamic. Distribution and archival standards will continue to be affected by technological change. For all the progress to date, we have yet to achieve fully error-free exchange of information, easy retrieval of archived data sets, or interoperability of hardware and software. More technological changes are inevitable. The size and sophistication of the user community, high-density storage media, and an unparalleled applications software development effort have rewritten rules for data standards. And, in the judgment of these authors, it is the combination of the three that has had the greatest impact.

The globalization of data uses necessitates communication on the issue of standards. International business, international academic research, and United Nations programs are examples of sophisticated uses of data sets requiring new distribution and archival standards. The integrated European market; the political changes in Germany, Eastern Europe, and the former U.S.S.R.; potential tariff agreement in the Western Hemisphere; and strengthened copyright laws in the Pacific Rim countries will accelerate the push for easy communication among data users and promote the development of international standards.

Hands-on Users
Without a common understanding of distribution and archival standards and principles, forthcoming changes may not all be improvements. As the largest producers of research data, government agencies worldwide face the unprecedented challenge of serving a rapidly growing end-user market now numbering in the millions. Today, data end-users number in the tens of millions and all of them expect to interact directly with the data product.

During this period of flux, the development of distribution and archival standards can benefit from input by a knowledgeable user community. And, IASSIST members are on the leading edge of understanding the need for governments and analysts alike to practice “safe data.” As the traditional technical role of the programmer has faded, end-users have acquired a new responsibility for safe-guarding the welfare of their own data resources and distribution channels. Yet, it is unclear how much of this responsibility users will co-opt for themselves and how much they will hand back to the professional data processing community.

What will be role of the data archivist as we move into the 1990s? How many of the distribution, retrieval, and archiving functions will be done by business professionals, research analysts, and computer professionals? IASSIST members will be facing these questions head-on in the coming years.
Selected Bibliography


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

2 Digital documents are defined to be those that reside, are distributed, or primarily archived in digital form — traditionally on magnetic tape.

3 The term archival is used throughout this paper to mean the transition or transformation of digital-data from that form normally associated with daily or regular use to an alternate compressed form intended for less frequent use and/or long-term historical storage.

4 MB is an abbreviation for mega-byte or million bytes (characters) of storage. To put this in context, an average single-spaced page of text contains approximately 3,200 characters. Thus, a 10MB hard disk could store the equivalent of approximately 3,100 pages of text.

5 The original 8086 operated at an internal clock speed of approximately 4.7 MHz (4.7 million cycles/second). A 350MB hard disk can hold the data traditionally stored on approximately 10 magnetic tapes recorded at 6250 BPI, the current recording density.

6 Scientific experiments now frequently incorporate instrumentation that measure such information as temperature which is automatically captured by PCs as the experiment occurs. Other programs running on PC analyze these data continuously as the experiment occurs.

7 CAI is an acronym for computer assisted interviewing. SPSS, SAS, BMD are all statistical packages that began on the mainframe and which are now available for use on the PC.

8 MS DOS is an acronym for Microsoft Disk Operating System. It is the dominant operating system (control program) used by microcomputers. The chief rival to MS DOS is the Apple MacIntosh.
Using The Century Of Prose Corpus

by Louis T. Milic
Cleveland State University

The Century of Prose Corpus (COPC) is one of a number of compilations of texts that have been developed during the last three decades to facilitate a certain kind of linguistic analysis with computers. Unlike the Corpus Thomisticus (the whole of the works of Thomas Aquinas), for example, the kind of corpus I am talking about is a descendant of the Brown Corpus, devised in 1961 by Henry Kucera and Nelson Francis of Brown University. The Brown Corpus consists of a million words of edited American prose, all published during that one year and taken from a great variety of kinds and genres of printed materials, from humorous fiction to articles in learned journals. The devisers assumed that their corpus was large enough to represent nearly every type of linguistic unit that might be of interest to scholars. Although it was intended to be machine-readable, it has generated two large volumes of analysis and documentation in which alphabetic and rank-ordered word-lists provide a view of the American vocabulary at that period, among many other valuable pieces of information about the language, to say nothing of the other areas of knowledge that are served by this work.

It would not be inaccurate to compare the Brown Corpus and other corpora that have sprung up since to anthologies, such as those that serve as textbooks in courses in literature, history and other fields. The anthology is more than anything else a sample of the writing of a field or period, representative and typical of the totality of the population. Of course, it is not a statistical sample because of its preference for the best, the best-known, the most influential..., but it is a sample nonetheless. Someone who has read through an anthology has a grip on the writing and thinking, the preoccupations of a genre, time period, nation... Similarly, anyone who had been living on another planet during 1961 and on his return read through the million words of the Brown Corpus would have a pretty complete idea of what went on during that year. But of course that was not the intention of Francis and Kucera; their compilation was primarily a tool for research in language. Their Corpus gave rise to similar ones of Spoken English and of British English. But beyond that, it led others to create more specialized corpora. The COPC is one such.

The COPC is intended to represent a norm for the study of the English of Britain during the eighteenth century. Its actual delimitations are the years 1680-1780 and its dimensions are approximately 500,000 words. It is composed of two parts: A) the major authors:

Addison, Berkeley, Bolingbroke, Boswell, Burke, Chesterfield, Defoe, Dryden, Fielding, Gibbon, Goldsmith, Hume, Johnson, Locke, Adam Smith, Smollett, Steele, Swift, Temple, Walpole;

B) the 100 background writers.

Part A (the major authors) contains 15,000 words from each of the twenty most prominent authors in three selections of 5,000 words each drawn from various stages of each author's production. This part totals 300,000 words or 60% of the Corpus. Part B can best be visualized as a ten by ten matrix, in one dimension representing decades of years:

1=1680-1689  2=1690-1699  3=1700-1709  4=1710-1719
5=1720-1729  6=1730-1739  7=1740-1749  8=1750-1759
9=1760-1769  0=1770-1779.
and in the other ten different genres:

<table>
<thead>
<tr>
<th>1 Biography (A)</th>
<th>6 History (G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Periodicals (B)</td>
<td>7 Memoirs/Letters (H)</td>
</tr>
<tr>
<td>3 Education (D)</td>
<td>8 Polemics (K)</td>
</tr>
<tr>
<td>4 Essays (E)</td>
<td>9 Science (N)</td>
</tr>
<tr>
<td>5 Fiction (F)</td>
<td>10 Travel (Q)</td>
</tr>
</tbody>
</table>

It will be noticed that there is a blending here of genre and subject matter, which can be rationalized by the claim that subject matter dictates conventions that amount to genre.

There is a text of 2000 words in each cell. Consequently there are ten selections (20,000 words) for each decade (one from each genre) and ten for each genre (one from each decade), the whole consisting of one hundred selections of 2000 words each or 200,000 in all, 40% of COPC.

Each sentence of each text is identified by means of a header block. An excerpt of one of the Part B texts follows:

```
5N03(1728)0001/021-P1 Language is a set of words which any people have agreed upon, in order to communicate their thoughts to each other.
5N03(1728)0002/079-P0 The first principles of all languages, Buffier observes, may be reduced to expressions signifying first the subject spoken of; secondly the thing affirmed of it; thirdly the circumstances of the one and the other: but as each language has its particular ways of expressing each of these; languages are only to be looked on as an assemblage of expressions, which chance or caprice has established among a certain people; just as we look on the mode of dressing, etc.
```

The header block 5N03(1728)0001/021-P1 is analyzed thus:

```
5N03: identifier of text (decade 5, genre N, accession no. 03)
1728: date of publication
0001: sentence number 1 of selection
021: number of words in sentence
P1: sentence begins a paragraph.
```

The entire Corpus holds on three high-density 3 1/2" diskettes (or on tape) and may soon be available on CD. It can be used on mainframes or on 386-type personal computers. In its present form, it requires the user to have access to a program package (such as EYEBALL, ARRAS, Word Cruncher...) or to be able to program in a string-manipulation language, such as SNOBOL or one of its derivatives (e.g., SPITBOL). I have devised several programs with which I have analyzed the various texts for later statistical treatment. I shall mention two of these.

The LETTER program performs the following:

1. Counts the length of each sentence
2. Produces a sequential list of sentence lengths
3. Calculates and prints
   a. mean sentence-length in words
   b. standard deviation of the sentence-length
4. Displays a frequency distribution of the letters in the text, both raw scores and percentage
5. Displays the rank-order of the letters according to frequency, compared to the Brown Corpus and other corpora
6. Displays a frequency distribution of word sizes
7. Displays frequency distributions of word-initial and word-final letters for words greater than five letters in length.

8. In a summary, provides the following:
   a. total words
   b. hyphenated words
   c. number of sentences
   d. number of interrogative sentences
   e. net number of letters in the text
   f. calculated vowel-consonant ratio
   g. mean sentence length
      1. in letters
      2. in words (by a method different from 1)
   h. mean word-length in letters.

The INDEXER program does the following:

1. Alphabetical word-index with raw frequencies
2. Rank-ordered word-index of the 100 most frequent lexemes, with raw frequencies and percentages
3. Counts
   a. tokens
   b. types
   c. hapax legomena
4. Calculates
   a. type-token ratio
   b. hapax-token ratio
   c. hapax-type ratio.

And of course, the programs may be applied not only to individual texts, but to groups, to decades, genres, Parts and to the whole corpus.

As can easily be noticed, these two programs alone acting on each of the texts in COPC generate a very substantial amount of data which can be analyzed or treated in a number of ways. To illustrate one possibility out of many, I shall follow Newton’s principle about the relation of data and hypotheses:

For the best and safest method of philosophizing seems to be, first diligently to investigate the properties of things and establish them by experiment, and then to seek hypotheses to explain them. For hypotheses ought to be fitted merely to explain the properties of things and not attempt to predetermine them...

Inspection of the data - that is, the texts themselves and the output of the programs - had led me to observe that writings in the same genre showed a consistent use of certain variables. In order to examine this possibility, I organized the data of Part B into ten variables for each of the hundred texts in it and analyzed this by means of the SPSS statistical data analysis package. Although the number of variables is of course unlimited, I chose ten more or less at random. These variables consist of two sets: “standard” and arbitrary. The five standard variables (often found in the literature):

1. mean sentence-length (MSL)
2. mean word-length (MWL)
3. number of types (TYP)
4. number of hapax (HAP)
5. percentage sum of five most frequent function words (FW)

The five arbitrary variables are:

6. frequency sum of the letters “t,” “i,” and “o” (LET).
7. frequency of the letter “s” in final position (SFIN).
8. frequency of the letter “d” in final position (DFIN).
9. sum of the two most frequent function words (TOP2).
10. number of nouns in ranks 1-54 of each selection (NN).

The Correlation procedure in SPSS produces Pearson correlation coefficients for each pair of variables, when these have been arrayed in an appropriate form, as follows:

<table>
<thead>
<tr>
<th>Text</th>
<th>MSL</th>
<th>MWL</th>
<th>LET</th>
<th>SFIN</th>
<th>DFIN</th>
<th>TYP</th>
<th>HAP</th>
<th>FW</th>
<th>TOP2</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>6Q45</td>
<td>29.82</td>
<td>4.48</td>
<td>24.17</td>
<td>22.20</td>
<td>12.64</td>
<td>695</td>
<td>455</td>
<td>21.64</td>
<td>13.85</td>
<td>1</td>
</tr>
<tr>
<td>8Q16</td>
<td>40.86</td>
<td>4.65</td>
<td>23.12</td>
<td>17.79</td>
<td>15.06</td>
<td>762</td>
<td>534</td>
<td>19.04</td>
<td>10.14</td>
<td>7</td>
</tr>
<tr>
<td>0B90</td>
<td>38.63</td>
<td>4.49</td>
<td>24.18</td>
<td>15.56</td>
<td>18.63</td>
<td>846</td>
<td>615</td>
<td>19.64</td>
<td>9.90</td>
<td>3</td>
</tr>
<tr>
<td>0N74</td>
<td>62.47</td>
<td>4.85</td>
<td>24.63</td>
<td>28.49</td>
<td>15.56</td>
<td>737</td>
<td>510</td>
<td>21.65</td>
<td>12.25</td>
<td>10</td>
</tr>
<tr>
<td>4D83</td>
<td>45.52</td>
<td>4.70</td>
<td>24.70</td>
<td>23.54</td>
<td>15.32</td>
<td>637</td>
<td>385</td>
<td>18.97</td>
<td>10.23</td>
<td>6</td>
</tr>
<tr>
<td>8K78</td>
<td>34.02</td>
<td>4.55</td>
<td>25.80</td>
<td>17.46</td>
<td>13.73</td>
<td>683</td>
<td>440</td>
<td>19.12</td>
<td>10.59</td>
<td>6</td>
</tr>
<tr>
<td>6K82</td>
<td>46.88</td>
<td>4.54</td>
<td>25.25</td>
<td>20.89</td>
<td>9.93</td>
<td>578</td>
<td>339</td>
<td>20.30</td>
<td>9.68</td>
<td>8</td>
</tr>
<tr>
<td>9B73</td>
<td>38.32</td>
<td>4.70</td>
<td>24.78</td>
<td>24.11</td>
<td>11.17</td>
<td>740</td>
<td>495</td>
<td>20.88</td>
<td>11.57</td>
<td>7</td>
</tr>
<tr>
<td>9K98</td>
<td>34.00</td>
<td>4.64</td>
<td>24.40</td>
<td>25.08</td>
<td>14.98</td>
<td>627</td>
<td>395</td>
<td>19.79</td>
<td>11.62</td>
<td>9</td>
</tr>
<tr>
<td>2N09</td>
<td>34.50</td>
<td>4.62</td>
<td>23.55</td>
<td>27.04</td>
<td>12.07</td>
<td>669</td>
<td>442</td>
<td>21.90</td>
<td>13.10</td>
<td>7</td>
</tr>
</tbody>
</table>

The Pearson correlations are as follows:

<table>
<thead>
<tr>
<th>Positive</th>
<th>.01</th>
<th>.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSL-FIN</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>MSL-FW</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>MWL-HAP</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>SFIN-TYP</td>
<td>.29</td>
<td>.46</td>
</tr>
<tr>
<td>SFIN-HAP</td>
<td>.28</td>
<td>.55</td>
</tr>
<tr>
<td>SFIN-FW</td>
<td>.24</td>
<td>.46</td>
</tr>
<tr>
<td>SFIN-TOP2</td>
<td>.29</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TYP-NN</td>
<td>-.28</td>
</tr>
<tr>
<td>HAP-NN</td>
<td>-.30</td>
</tr>
<tr>
<td>LET-TYP</td>
<td>-.31</td>
</tr>
<tr>
<td>LET-HAP</td>
<td>-.29</td>
</tr>
</tbody>
</table>

As can be easily seen, a good number of these are quite significant, some at the one percent, some at the next level, mostly positive, although a few are negative. Of course, some of the correlations are significant but meaningless, as they represent merely functional relationships, e.g., types and hapax, function words and “top two”. But others suggest something factual and possibly important about the relationship of genre to the quantitative fabric of texts. To look into the possibilities of this relationship, we must go deeper and discover which genres select which variables. By subjecting the data to analysis of variance (ANOVA), we find the pattern in the following matrix:


It is plain that certain genres select more significant variables than do others. Numbers 4 and 7 (essays and memoirs/letters) seem less distinct than the others. Numbers 5 and 9, on the other hand (fiction and science), are much more distinctive.

Following Newton’s recommendation, therefore, we are free next either to devise hypotheses about these relationships or try new experiments to deepen our understanding. A possible explanation might be that the conventions of fiction and science writing are much more strict than those of essays, memoirs or letters, and that this strictness manifests itself at the quantitative microlinguistic level. Another might be that the term “genre” is not as rigorous or as easily defined as is generally believed. At any rate, to feel confident about such hypotheses would require further analysis of factors by means of regression or other advanced statistical techniques.

This simple illustration is only intended to reveal a small fraction of the immense possibilities for study and research that are latent in a carefully constructed corpus of substantial size and extent.

1 Presented at the IASSIST 92 Conference held in Madison, Wisconsin, U.S.A. May 26 - 29, 1992.
Exchange of scanned documentation between social scientists and data archives: establishing an image file format and method of transfer

by Repke de Vries and Cor van der Meer
Steinmetz Data Archive for the Social Sciences
Amsterdam, Holland

Introduction
Social science research uses as its raw material not only datasets but also the accompanying documentation: codebooks, questionnaires and so on. Sometimes these “guidebooks” are machine readable and available as text files. But older studies and questionnaires in their original form are all paper - only documentation. Other examples are handwritten comments on computer print out, sketches and black and white pictures as used in psychological research. Needing this kind of documentation means repeated photocopying by archive or library and mail delivery whereas the actual data may travel by networks like the Internet or be put on tape or any other computer medium. It's a situation disadvantageous to both the archiving world and the researcher in need of complementary documentation - especially if both are geographically wide apart.

Wasn't the Fax machine invented to do just that - to get any sketch, image or piece of text instantaneously from A to B? To an extent yes - but the resolution is poor and it is still repeated “photocopying” sending and lots of paper again upon receiving. The image can't be pasted in a research paper, nor can it be stored in a database, viewed on screen or read by OCR packages. Fax boards and a personal computer don't change that really: for one thing there can't be constant polling for incoming Faxes or a direct telephone connection is not available to the researcher. And though a Fax board gives you the image (whatever it is) for the first time as a file on the PC, the resolution is still not good enough.

Networking on the other hand is mature now: the integration of local area networks with interconnecting nets like the Internet, often gives the desktop computer global networking facilities whereas the one Fax machine for the department is down the corridor.

Obviously transferring codebook pages etc. as images has to follow a different scenario, avoiding the repetition and manual labour in Fax and taking advantage of network capabilities:

The scanning of the document has to be separate from transfer. Scanning should be a one time operation with adequate resolution. Storage has to involve compression techniques. The collection of image files could be handled by a specialised database that also holds descriptive and administrative information. Or the files might be the result of just scanning a few questionnaire pages with hand written comments. The advantage over Fax is that once scanned and stored, sending out an image - like any other computer file - is easily repeated and initiated. And such scanning can be done at a much higher resolution.

Storage formats for scanned images can be the own policy of archive or library but an exchange format (and the necessary conversion ) should be accepted and adhered to by anyone offering documentation as image files.

The transfer comes next and can be done in a number of ways, even as ordinary mail by reprinting the image on paper with a laser printer. Network transfer though is easiest and fastest. The researcher needing the pages receives it as a series of small files on his or her own computer or personal file area in a local network.

The last step involves a tool for the end user to decompress and actually use the images. Ideally the images received can be handled as such by the usual word processing software available to social science researchers. But a free software program will otherwise translate back from the exchange format to a "common denominator" format, if need be.

Establishing an Exchange Standard for images.
TIFF as the format of choice for the Exchange of Images.

The "tagged information file format" was launched by Aldus Corporation and Microsoft in 1986 and Revision 6.0 is now (April 1992) in Draft 2 and finalizing.

All this time careful attention has been paid to keep the skeleton of the TIFF header and the mechanism of the format (a pointer structure) the same. Older TIFF readers or writers therefore can exit gracefully if confronted with a TIFF file holding a state of the art colour image. Another feature is the use of tags holding vital information about the kind of image, the compression type used for the image block inside the TIFF file, but also texts of possibly any length describing the image.
If software can't read "TIFF" though it promises clearly to do so, it is just because of this versatility. Often simpler compression types possible in TIFF together with black and white images are handled but grey scale or a more complicated compression method are not. Reading appropriate tags in the TIFF file these packages could have given you helpful hints why it was decided that your TIFF variation can't be imported, but most of the time a misleading message on the screen mutters about "incompatible format". If one knows how to read the information, similarly a TIFF header dumper program tells you straight away how the image in the TIFF file is built up.

For data archives and libraries starting the service of making documentation available as images, it is of paramount importance to choose a standard that:

- has wide acceptance,
- is not in any way patented or licensed (with concern to the compression schemes),
- is not computer type or operating system dependent,
- has features to make it self-explaining (documentation tags)
- and is open to new developments in the imaging field but will never be changed in its basic format.

An indication of the acceptance of TIFF as standard for an image file format is the publication last January of the Memo "A file format for the exchange of images in the Internet" by the Network Fax Working Group of the Internet Engineering Task Force. Authors Alan Katz and Danny Cohen from USC Information Sciences Institute, define "the standard file format for the exchange of bitmapped images within the Internet" as a particular TIFF variation. (TIFF-B, preferably with compression type 4).

TIFF is the format read without any problem by the major OCR programs. Format stability is an issue close to the heart of archives. For the storage of images the long term perspective is carefully planned for in the development of the TIFF standard. On the other hand the TIFF 6.0 Revision draft also shows how flexible the standard really is: if libraries or archives take an interest in offering photographic information as images, the same TIFF format can act as wrapper but this time with JPEG compression that is now accepted as one of the TIFF compacting schemes.

In choosing the right kind of TIFF format for the Exchange standard, the following is presumed:

- foremost is the need for scanning and transferring of text together with some line drawings, as in questionnaires. These are called black and white or, bilevel images.
- the scanning resolution should be 300 dpi. This gives adequate detail and matches best with the printing resolution of today's average laserprinter. Mismatches complicate the software needed to either convert to the exchange standard or use the images afterwards.
- the compression chosen should be optimized for bilevel images and pack as tightly as possible
- each original page of information is kept as separate image and separate TIFF file; TIFF has a multi-page feature (one resulting file, holding a number of compressed images) but this option is for the moment not used.
- there is a need for adding descriptive information to the image; TIFF has tags that can be used for that purpose but this option is for the moment not used.

This leads to the choice of TIFF compression type 4. Well described in the TIFF 5.0 paper, still present in the TIFF 6.0 Revision draft (draft 1, February 1992) as one of the compression schemes. This compression type follows Fax Group 4. (The two numbers "four" are a coincidence). And Fax Group 4 is yet another standard and already fully described in the CCITT Recommendation T.6. The compression and decompression techniques described in the Recommendation are open to anybody for use in own programming. Fax Group 4 is optimized for bilevel images that hold a mix of text and lines: a lot of white with interspersed black dots.

The tags used are (referring to TIFF revision 6.0, February 1992): the Architectural fields and the Resolution fields, both Baseline TIFF fields. TIFF 6.0 has paragraphs in "Section 4" (another coincidence) that further define these fields given bi-level images. Note that in the text mentioned, compression type 2 is used as a working example whereas the Exchange standard employs type 4.

In the future the TIFF Informational fields and Document Storage and Retrieval tags could be exploited to make images self explaining. Contrary to the (unused) multi-page feature of TIFF, these fields or tags can be handled and inspected by the user with any common file viewer: the information stands out as readable text among garbish (though this garbish is a Sleeping Beauty: it is the scanned and compressed image). Either direct at the beginning of the TIFF file or at the very end.

The Steinmetz Archive will help with all necessary
documentation and expertise if a data archive or library wishes to implement it's own TIFF compression type 4 writer or reader. The Archive will also provide a testbank service to judge if one starts using the Exchange Standard with indeed the right TIFF 4 format.

Further reading, commercial conversion packages, shareware TIFF viewers/printers and the anonymous FTP availability of the excellent "Sam Leffler" toolkit to start programming for TIFF - are mentioned at the end of the paper.

The Katz and Cohen proposal, also defines for bilevel images but is less strict in TIFF compression type and resolution of the scanned images. The Working Group allows even uncompressed TIFF for example, though TIFF type 4 is to be preferred. Multi-page files are supported. The perspective however of the proposal seems different from ours: Katz and Cohen have a strong emphasis on the actual transfer of images and leave it to the sending and receiving parties to negotiate a variation of their standard that both can handle. Our emphasis is on establishing an Exchange standard that ensures the researcher that he or she can always use the images received. Hence one compression type and so on.

Producing images by document scanning.

Given the nature of printed text and line art, scanning black and white at 300 dpi or more is adequate. Issues of preserving grays in the original or even colour are not involved. Each separate page scans into one compressed file and these files are kept together by proper file names and subdirectories or folders to mimic the original chapters and separate volumes.

Especially in a closed system scanning station with proprietary software it is not always made clear by the supplier how the images are stored in terms of image format and compression type. The compression (decompression) more over is often done by additional, separate hardware. In order to exchange it is imperative that the system has exporting facilities so that images can be converted to an established standard. Next these converted images should be available in a general file area, open to networking and further handling.

Scanning and storing in a Dos environment without specialized image bank software is more or less open by definition and can produce accessible images in a TIFF format straight away, though often the less compressing TIFF type 2 or even TIFF Packbits is used.

Both closed and open approach don't necessarily produce the TIFF type 4 chosen as exchange format straight away. The following steps have to be taken:

First case: a scanning station with own image database software and hardware compression and decompression. Used for systematically scanning all paperwork of a number of studies. If there is a choice at all and if one only scans bilevel (black and white) printed source, TIFF type 4 is a very good choice for an internal storage format as well. If the software is custom made, even the TIFF documentation tags can be filled in to make the image files self explaining.

Second case: if the scanning station comes as is, caveat emperor:

- given your computing environment, the image files should still be open for access by other software

- if internal format "type X" is used, this format should be convertible by both software and hardware to the required TIFF compression type 4, Exchange format. "Hardware" means that the scanning station software asks its compression/decompression board to do the conversion. But can it? "Software" means that a separate tool is available or can be written to convert to TIFF type 4. If necessary both approaches can be split in successive steps: if only TIFF 2 or 3 can be managed than an off the shelf graphics format conversion package, can do the TIFF 2 or 3 to TIFF 4 step. Note that a scanning setup that uses a storage format that depends entirely on separate hardware, is a timebomb for a data archive or library. At some point in time the hardware board will fail and if a replacement is no longer available, the whole collection of scanned images is rendered useless.

- if TIFF is used as internal format (and given bilevel images) it is very likely to be TIFF type 4, because it is the most suitable compression scheme. If not than probably a standard package can do the conversion to the TIFF 4 Exchange format.

Third case: a simple scanner setup with some software for viewing and image manipulation, attached to a PC and used for per request scanning of documentary information.

All involved packages in this case handle TIFF but only aiming at import into desktop publishing software, of the wrong, simpler compression types. Standard conversion packages can change into the required TIFF 4.

Note: it is desirable to use this most compact TIFF 4 format also for storage to accommodate future similar requests.

Transferring a group of images.
One of the features of the TIFF format is storing several different images (pages of text) in one file. Scanning and storing however is done on a one page, one image basis. Therefore extra processing would be needed to use this feature and it does not really improve the transfer. Many smaller files travel easier over a net than one big chunk and dissecting and decompressing would be more complicated for the end user. Consequently this feature is not yet part of the Exchange standard.

Compressed images are binary files so in sending over a net, care should be taken to use the transfer protocols accordingly. If the facility is Text oriented - like Send File in BITNET, extra steps are necessary, (uencode and uudecode) to nevertheless preserve the binary nature of images. The FTP protocol used in Internet, offers both binary transfer and multiple put to handle a stream of images with one command.

Name giving of the image files can be a problem: different operating systems have different conventions and too long a name gives trouble if the receiving end has a simpler scheme. The best seems the DOS convention (8 characters, dot and a three character extension) just because it is the most restricted one and will fit into any other notation. Unfortunately this means that if TIFF type 4 is used for internal storage as well and the platform is Unix with rich name giving possibilities, one still needs a conversion of the file names. This can be done while copying from the image storage area to the transfer area.

The user side: transforming back the images to screen, paper or OCR file.

Implementing an Exchange standard, the focus of attention should of course be the ease of operation for the user to wave the magic stick and have the requested documentation on screen or on laserjet printout. If the data archives or libraries offering the service take the trouble of converting to one and the same exchange format (TIFF 4, 300 dpi, single page per file), the steps to be taken are well defined. If text processing software handles graphics, it can import TIFF (but only the simpler compression types) and print it out. Drawing or imaging software does the same and offers viewing. Disadvantage of this approach are the required expertise to import a received image into one’s word processor and above all: get it printed. Drawing software is specialized in importing, viewing and printing images but certainly not everybody masters that kind of software. For various platforms good shareware software is available that reads TIFF (again: the simpler compression types) and lets you both view and print. All software requires a setup of for the DOS environment - 286 or 386 PC with VGA and a laserprinter available. This printer should be equipped to also print larger chunks of “graphics”: an image holding a full page of text is a bit too much for laserprinters with limited graphics capabilities.

A few pages of “how to do” information for some software common to most researchers, could ease this Do It Yourself approach to use tools already available. Such a document will be made available by the Steinmetz Archive and will also point out the usefulness of some shareware already specialized in doing the job.

Remains the demand by popular software to be on a “simple compression TIFF diet”. Clearly a conversion aid is needed to change TIFF compression type 4 into one of the simpler schemes mentioned. The Steinmetz Archive has written a tool to do just that and will make it available to data archives to bundle it with requested images. Ideally this tool will also have the option to print the decompressed image directly to a HP LaserJet printer. Printing is much easier accomplished than viewing because of the wide variety in display hardware and the limited resolution or viewing area of most PC screens. The HP printing option will most certainly be included in a future update by the Steinmetz Archive. (Postscript printing would be desirable).

With a growing user base for TIFF type 4 bilevel images, software makers can be urged to implement Importing and handling this TIFF compression scheme as well. Then the separate conversion step is no longer necessary. In the area of OCR programs this already is the case: tests showed that the market leading OCR programs for DOS read the Exchange standard format without need for conversion. (Recognita, OmnPage, Prolector, Lioocr, K5200)

Summing up, these are the steps for the user to transfer back:

1. use the free conversion tool to simplify the TIFF compression type

2. with the help of the Cookbook print or view the images by applying existing software: either commercial packages or shareware. The choice should be software commonly available to social science researchers. (The shareware can be redistributed together with the Cookbook - both as files through networking and electronic mail)

3. use the Exchange format without further ado (for example with OCR software)

Further reading, availability of software and source
"A file format for the exchange of images in the Internet" by the Network Fax Working group of the Internet Engineering Taskforce. Authors: Alan Katz (Katz at ISI.Edu) and Danny Cohen (Cohen at ISI.Edu). Phone: 310-822-1511.

The Sam Leffler TIFF toolkit:
by anonymous FTP:
sgi.com: /graphics/tiff/v3.0.tar.Z (192.48.153.1)
email: sam at sgi.com
(This toolkit also includes the TIFF 6.0 specification)

Aldus can be reached at CompuServe, again the TIFF spec's and a much simpler TIFFRead toolkit

Commercial conversion packages to and from TIFF type 4:
(DOS) HiJaak
(DOS and SUN OS): Image Alchemy
(UUCP: hsi at netcom.COM or:
apple!netcom!hsi)

(DOS) Shareware graphics viewers and printers:
among others:
Graphics Workshop
Optiks (all three don't handle TIFF 4 so need the free conversion tool first)
Pixfolio (runs in a Dos Windows environment)

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

This paper was also presented at CSS92, May 1992, Ann Arbor, Michigan.
The Wisconsin Longitudinal Study: Adults As Parents And Children At Age 50

by Robert M. Hauser 1, William H. Sewell, John A. Logan, Taissa S. Hauser, Carol Ryff, Avshalom Caspi, and Maurice M. MacDonald, Institute on Aging and Adult Life and Center for Demography and Ecology The University of Wisconsin-Madison

Summary
We are carrying out a survey of more than 9000 American men and women who were first interviewed as seniors in Wisconsin high schools in 1957 and have subsequently been followed up in 1958, 1964, and 1975; these individuals will be approximately 53 years old when they are interviewed in 1992. At the same time, we will interview a randomly selected sibling of most respondents. Approximately 2000 of these siblings were previously interviewed in 1977, and we have sufficient resources to interview approximately 4000 more siblings during this round of the study. The data collection process will include a 1-hour telephone interview, followed by a self-administered mail questionnaire, and a waiver that will permit us to link our survey records to information from the Social Security system. Finally, we expect to obtain enough information to link our records to the National Death Index.

Data from the Wisconsin Longitudinal Study (WLS) will be a valuable public resource for studies of aging and the life course, inter-generational transfers and relationships, family functioning, social stratification, physical and mental well-being, and mortality. The study has 5 specific goals: (1) To extend models of occupation and earnings and to elaborate the roles of aspirations in adolescence and at mid-life, of previous achievements, and of familial responsibilities in current economic and social standing, subjective well-being, mental and physical health, disability, and wealth; (2) To identify and measure local effects on opportunity, that is, specific characteristics of a person, firm, or economic sector that directly influence the chances of obtaining a job or a limited range of jobs; (3) To extend and elaborate models of sibling resemblance that will elucidate influences of the family of origin on the life course; (4) To investigate self-assessments of well-being in the context of aspirations, accomplishments, and social relationships with significant others; (5) To measure social and economic exchange relationships with parents, children, and siblings and assess the consequences of those relationships for well-being.

We are planning a follow-up survey of more than 9000 American men and women who were first interviewed as
and children, and in the context of past and current social relationships with those significant others; (5) To measure social and economic exchange relationships with parents, children, and siblings and assess the consequences of those relationships for well-being.

Background And Significance
The Wisconsin Longitudinal Study (WLS) is a long-term study of a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957. Survey data were collected from the original respondents or their parents in 1957, 1964, and 1975. These data provide a full record of social background, youthful aspirations, schooling, military service, family formation, labor market experiences, and social participation of the original respondents. In 1977 the study design was expanded with the collection of parallel interview data for a highly stratified subsample of 2000 siblings of the primary respondents.

The WLS Data
The WLS is a rich source of data on life-cycle processes that is of continuing interest to scholars in sociology, education, psychology, and economics. The interview data have been supplemented by mental ability tests of primary respondents and siblings, measures of school performance, and characteristics of communities of residence, schools and colleges, employers, and industries. The WLS records for primary respondents are also linked to those of three, same-sex high school friends within the study population. The measurement of social background includes earnings histories of parents obtained from Wisconsin state tax records, and the data on the socioeconomic careers of men in the main sample are supplemented by social security earnings histories from 1957 through 1971. The WLS is widely recognized as one of the most useful bodies of longitudinal data on the lives of Americans because of the quality of the survey measurements (and our efforts to measure that quality), extremely high retention of panel members, complete, multi-layered documentation of the data, and multiple linkages to personal and institutional records.

Research Based on the WLS
The WLS panel has been used to develop the well-known "Wisconsin Model" of social and psychological factors in socioeconomic achievement. We have located more than 800 SSCI Citations to 7 core WLS publications since 1972. In addition, or in extensions of this central line of research, the WLS data have been used in studies of geographic constraints on college access; recruitment into teaching, nursing, and other occupations; choice of marital partner; differential family formation and fertility; gender differences in market participation and success; religious and ethnic differences in achievement processes; birth order effects on ability and achievement; effects of high schools and colleges on aspirations and achievements; and inter-firm and inter-industry differences in compensation. Also, the project has been the locus of many useful methodological developments built around the design, collection, or analysis of data from the WLS. These include successful methods for tracing respondents over long intervals; the analysis of unit record data from the Social Security Administration, and the usage of data on other socioeconomic variables and models of common family factors in the achievements of siblings 6

Our last direct contact with the primary WLS respondents took place in 1975, when they were about 36 years old. At that time, most of the women were completing childbearing and were participating in the labor market or planning a return to it; men were well established in their occupational careers, but - because they married younger women - were not as far along in family formation. Using these data, we have analyzed the process of socioeconomic achievement from adolescence to midlife and compared the socioeconomic achievement processes of men and women. WLS siblings varied widely in age, but 80 percent were between 27 and 45 years old in 1975, and for adult sibling pairs, we were able to conduct studies of family resemblance and intrafamily differences in education, occupation, earnings, and fertility. Among our main research goals are to extend our models of social and economic achievement and participation of primary respondents and their siblings.

Planned Follow-Up Surveys
In summer 1992 we began to interview the 9000 primary respondents and, whenever possible, a randomly selected sibling of each. The primary respondents will be 53 years old, and four fifths of their siblings will be 44 to 62 years old. At those ages, the WLS respondents and their siblings will be anticipating their own retirement and aging as well as managing relationships with one another, their adult children and their elderly parents: (1) In 1975, 92 percent of respondents had at least one living sibling, and 71 percent had at least two. Moreover, because of their position at the leading edge of the baby boom, siblings tended to be younger than primary respondents. Thus, we expect that an overwhelming majority of WLS respondents will still have at least one living sibling. (2) In 1975, 93 percent of the respondents had at least one living child; since child-bearing began around age 18 (for women) and was not yet complete in the cohort, we expect that almost all respondents will have at least one adult child. (3) Survivorship is much less among the respondents' parents. We ascertained the
father’s year of birth in 1975, and we estimated that 18 percent of respondents will have a living father in 1992, while 42 percent of respondents will have a living mother. We estimate that about half the respondents will have at least one living parent, and the age of these parents will be around 80 years. Thus, we believe that our respondents are ideally suited for a study of aging and of intergenerational relations among adults.

In our 1992 interviews, we are updating our measurements of marriage and divorce, child-rearing, education, labor force participation, jobs and occupations, social participation, and future aspirations and plans among primary respondents and siblings. In addition, we are expanding the content of the study by obtaining data about psychological well-being, mental and physical health, wealth, and social and exchange relationships with parents, siblings, and children. In designing the new measurements, we have attempted to maintain an appropriate balance between comparability with our own previous concepts and methods (which are similar to those used in the Current Population Survey and the 1973 Occupational Changes in a Generation Survey) and comparability with other significant research efforts, e.g., the new Survey of Health and Retirement, the National Survey of Families and Households, NIH surveys of work and psychological functioning, and the NORC General Social Survey; in addition, we have coordinated our design efforts with those of members of the MacArthur Foundation Research Network on Successful Midlife Development. Finally, we plan to obtain information and waivers that will eventually link our survey data to Social Security records and the National Death Index.

We have considered whether the collection of new data for the WLS is warranted, given the existence of other longitudinal studies and the possibility of collecting similar data for a new national sample. The latter alternative may be desirable for some purposes, but it would be most difficult, and probably impossible, to provide the wealth of background and life history data that are available from the WLS or other longitudinal studies. The more serious question is whether the WLS is worth further investment, relative to other longitudinal studies of similar vintage. We think it is, for several reasons: (1) The WLS data on the life course are unique in richness and quality. (2) Major national longitudinal studies that began in youth cover more recent cohorts. These cohorts are of interest in their own right, but none has reached the pre-retirement years. For example, members of the National Longitudinal Study of 1972 will be around 38 years old in 1992, and there are currently no resources for further follow-up activities. Those in the HSB samples of 1980 and 1982 will be 26 to 28 years old in 1992; members of the two younger panels in the 1967-68 National Longitudinal Studies of Labor Market Experience will be 40 to 50 years old in 1992; the oldest cohorts covered in the Monitoring the Future Surveys will be about 35 years old in 1992. (3) Other longitudinal studies are restricted in similar ways to the WLS, which covers high school graduates from Wisconsin, almost all of whom are white. For example, members of the Career Development Study were juniors or seniors in the State of Washington in 1965-66, and they will be about 43 years old in 1992. The members of the NORC survey of 1961 college graduates are essentially the same in age as those in the WLS, but the sample is substantially more restricted with respect to educational attainment. (4) Project Talent may provide a national sample that is just 3 years younger than the WLS. However, it lacks the linkages of the WLS to socioeconomic data, and there have been serious problems of sample coverage and data access throughout the history of Project Talent.

New Directions for Research in the WLS

We have considered several ways in which the research agenda of the WLS could be extended. We have decided to focus on three of these opportunities in our initial work, without foreclosing the development of others at a future date. We believe that each of these is scientifically important and that they are complementary to the design and content of the WLS: (1) Effects of special preferences, skills, and attachments; (2) mental and physical health at midlife; and (3) social and economic exchanges and well-being.

Local Effects

Most of the previous analyses of the WLS data have used continuous measures of outcomes—particularly, years of education, occupational status, and earnings—as dependent variables in structural equation models. This has improved our understanding of the relationships of a number of background and social psychological variables to education, occupation and earnings. However, the amount of variance explained by these linear models has always been relatively modest. This has been attributed to the operation of “luck” in individual outcomes (Jencks et al. 1972), but it may also arise from a systematic neglect of factors that are not easily captured by linear models. The planned new wave of data collection will attempt to measure persistent effects of some of these factors, called “local” effects.

A local effect on occupational opportunity is any characteristic of a person or of a firm or economic sector which directly influences the chances of obtaining only a limited range of jobs. It is contrasted with a “general” effect, such as the effect of general education, which influences chances of employment in a wide range of jobs. An example of a local effect would be a particular skill or aptitude, such as mechanical aptitude. There are some jobs, mostly in the middle range of prestige and income, which demand high mechanical aptitude. Possession of this aptitude should have a local effect on
an individual's occupational chances, raising the probability of landing the jobs requiring it, but not raising the probability of landing good jobs in general.

Aside from specific skills or aptitudes, three other main types of local effects can be distinguished, namely, preferences, contacts and structural shifts: (1) Individuals may, for reasons subject to empirical study, have preferences for certain kinds of work, such as outdoors jobs, jobs with less than usual amounts of direct supervision, or jobs with high creative or artistic potential. Jencks, Perman and Rainwater (1988) have examined non-monetary, non-prestige attributes of jobs and found them highly predictive of individuals' reports of job satisfaction, yet only poorly related to demographic measures such as age, sex, and education. We want to examine the relationship of such non-monetary, non-prestige preferences to particular life course developments, rather than to the measures just named. For example, preferences for certain job characteristics may vary with inter- and intra-generational family responsibilities, and stage of the life course. (2) Individuals may have direct or indirect personal contacts among those making hiring decisions in certain jobs. The desire of a parent to pass along a business to a child, the preference of a union for enrolling the children of its members, and the general social contacts of a parent or child which may be useful job leads for the child are all examples. Appropriate methods, which we expect to apply and refine, will make it possible to estimate the magnitude of these social network effects in a well-defined, general population. (3) Finally, the economy as a whole may experience contractions or expansions of opportunity in certain types of work. Such structural shifts cause transitory increases or decreases of opportunity in limited ranges of jobs. We are asking for retrospective descriptions of the first and last occupations held by respondents in their first two and last two businesses or organizations where each respondent has worked since 1975; in most cases, this will give us a complete employment history. Thus our occupational data will cover years with widely different levels of overall economic activity. It is important that models of local effects in occupational outcomes are not confounded with local structural effects; the broad temporal scope is intended to aid in distinguishing the two.

To put the overall point most simply, measuring and modeling "local effects" may explain more of the variation in occupational (and related) outcomes than can be done with regression, and may increase the qualitative detail of the explanations associated with multivariate studies of life course achievement in general populations. As individuals age and experience changes in their priorities and responsibilities in the post-childrearing, pre-retirement years of their fifties, qualitative aspects of the choices they make — as reflected in local effects — may produce more concrete explanations of behavior.

Mental and Physical Health

We plan to examine the influence of educational and occupational pursuits on mental and physical health. The inclusion of detailed measures of psychological and physical functioning to the telephone and mail instruments will strengthen the multidisciplinary significance of the WLS by linking the attainment process, typically the domain of sociology, to mental and physical health, typically the domain of psychology. The proposed linkage affords significant strides in several research domains. First, prior studies of well-being have documented connections with education and income for men and women in American society (Diener 1984; Veroff, Douvan, and Kulka 1981), but the effects have been small. However, previous studies have used single-item measures of well-being that are of questionable reliability and validity (Larsen, Diener, and Emmons 1985). These measures have shown no connection to theories of psychological health (Coan 1977; Jahoda 1958; Lawton 1984; Ryff 1989a) nor to related empirical measures (Ryff 1989b). The WLS employs a differentiated, multifactorial concept of positive functioning that incorporates not only global happiness and satisfaction, but also the respondents' assessments of their effectiveness in dealing with the external world (autonomy, environmental mastery), and their sense of direction and progress in life (purpose in life, personal growth). With additional measures of physical health status, it will thus be possible to map the effects of educational and occupational attainment on an array of mental and physical outcomes.

Previous research on the relation of social structural factors (e.g., education, income) to psychological functioning has also been largely descriptive. Previous studies chart the magnitude and direction of linkages between demographic characteristics and subjective well-being, but do not specify the mechanisms through which educational and occupational achievements affect self-evaluations.

Two central social-psychological mechanisms will be explored in the research: (1) We will examine how social comparisons with significant others influence subjective well-being in midlife. The parallel sibling sample in the WLS provides vital comparative data about the respondents' attainments relative to a key group of significant others. This question constitutes a significant departure from prior psychological research on siblings, which has focused on effects of sibship variables (e.g., number of siblings, birth order) on achievement, intelligence, and personality (Zajonc 1976), as well as on disaggregating the comparative effects of genetic and environmental factors on behavior (Plomin and Daniels 1987). Few studies have examined the nature of sibling relationships in adulthood and later life (Cicirelli 1989) or the consequences of these relationships for psycho-
logical well-being. It is likely that adults use their siblings as “measuring sticks” to evaluate their lot in life (Troll 1975). The WLS thus provides a compelling data set with which to study the influence of sibling relationships — and their inherent social-comparative features — on self-evaluations, subjective well-being, and physical health in midlife. The specific cognitive mechanisms through which such comparisons influence well-being are derived from a synthesis of relative deprivation theory (Suls 1986), Tesser’s (1988) self-evaluation maintenance model, and various strands of attribution theory (Mirowsky and Ross 1990). Additional comparative data will be obtained on the attentions of the respondents’ parents and children. Adults who have accomplished less than their parents may be at greater risk for psychological distress. Alternatively, the “American dream” suggests that parents hope to have children who do at least as well, if not better, than themselves, so negative discrepancies with children (i.e., when children have accomplished more) may be conducive to positive self-evaluations. These expanded self-other comparisons offer important new directions to research on intergenerational relations, which has neglected the midlife era when one’s children are becoming young adults and one’s parents are growing old (Hagestad 1987).

The second proposed social-psychological mechanism through which educational and occupational attainments influence self-evaluations is temporal comparisons. Those who have advanced considerably beyond their starting resources are expected to show more positive self-evaluations than individuals who have made little gain or have lost ground. It may not be absolute levels of education, income, or status that predict psychological well-being, but the magnitude of those attainments relative to the resources with which one began. The WLS provides significant advances over prior studies because we can operationalize temporal comparisons in a behavioral, performance-oriented domain (educational and occupational achievement). Previous research has examined only subjective perceptions of personality change (Markus and Nurius 1986).

In sum, the planned study combines a theory-guided view of psychological well-being with fresh ideas about the relevant social-psychological processes by which people evaluate their accomplishments within the context of enduring family bonds across the life course. The design weaves data on three generations and multiple siblings, enabling us to explore the dynamics of individual development in the context of family histories and generational succession.

Social and Economic Exchanges
Eggebeen and Hogan (1990, p. 4) have nicely stated the case for improved measurements of social and economic exchanges among parents and children: “In small-family societies, ... [t]heory thus suggests that parental investment will be diluted when a large number of children compete for resources, and that it will be more heavily concentrated on children who bear them grandchildren. ... These hypotheses have been difficult to evaluate for modern societies because of the paucity of data documenting patterns of exchange.” Our new data will include measures of exchanges including patterns of kin contact, financial assistance, and the provision of services and care-giving. In the context of extensive WLS information on family origins, marital and fertility histories, earnings records, and status attainments, measuring these exchange variables should permit major advances in our ability to test a variety of hypotheses about inter-generational transfers.

Following the family sociology tradition of Adams (1968), recent findings from the National Survey of Families and Households (Sweet, Bumpass, and Call 1988) have again demonstrated that patterns of kin contact, care-giving, and financial support tend to be intertwined. Relatives who help each other with one type of assistance tend to provide the others as well, and to communicate more frequently in person and via mail and telephone contacts. Although the most intensive care-giving assistance is provided for severely ill or disabled persons, help in the form of child care remains very important despite the trend toward purchasing that care on the market.

As opposed to receiving aid, giving follows a U-shaped pattern by age, and in a manner that is particularly salient for persons of the WLS sample’s ages: young and elderly adults get more aid, and middle-aged adults are much more likely to provide aid (Lee 1979; Morgan 1982). Additionally the female members of the sample are more likely to exchange services with their kin, with males involved more in financial exchanges (Eggebeen and Hogan 1990).

Along with the likelihood that WLS members are heavily involved in family transfers and exchanges because of their stage in the life-cycle and the interaction effects of gender and age, their social exchanges should also tell us more about the impact of several recent trends. These include increased female labor force participation, rising divorce rates, increased demands for government spending on programs for children, and the slow growth in wages since the early 1970s. Furthermore, the extent to which prime-age parents may have been able to offset the effects of these influences on their children may have been restricted by their own economic and personal difficulties, as well as by the need to plan for new obligations that arise from increased life expectancy for themselves and their parents.

The current sources of family financial support for WLS sample members will include gifts and loans from older
parents and other relatives, as well as actual and expected bequests. However, many WLS members are likely to donate substantial time and financial support to their parents, as well as to their own offspring. Also, donations by grandparents to the children of WLS respondents may alleviate financial pressures for some. Economists who have studied these inter-household transfers (ITs) emphasize them as potentially important determinants of economic status that may have substantial redistributive effects (Cox and Raines 1985; Kurz 1984). Furthermore, the literature about the effect of Social Security on savings and retirement behavior has long recognized the potential of ITs to either complement (Cox 1987) or offset income opportunities from public transfer programs (Barro 1974, Lampman and Smeejing 1983).

Whatever their effect on the mix of support from family and public sources, it is clear that if ITs substantially augment the resources available to pre-retirees, they are likely to affect work behavior via wealth effects on labor supply and savings decisions (Kotlikoff 1987). Consequently, there is a need to identify which WLS respondents receive substantial ITs, to improve understanding about their role as a potentially important reason for heterogeneity in work behavior, social exchanges with kin, and psychological well-being. Additionally, analyses of the circumstances that motivate WLS respondents to donate substantial ITs to their parents, children and siblings can help to elucidate how the financial pressures of those responsibilities influence earnings and other economic status variables.

Previous Research

Previous research with the WLS developed comprehensive social psychological models of socioeconomic achievement from adolescence through age 36. In recent work, our aim has been to incorporate estimates of response errors in variables entering into the models and to account analytically for similarities and differences between siblings. We studied the extent to which the parameters of our stratification models were distorted by random and correlated errors in reports of parental status, social influences, and educational and occupational aspirations and attainments. We incorporated our estimates of errors in these variables into attainment models both for men and for women. We have made considerable progress in our research on sibling similarities and differences in socioeconomic careers and in family formation and fertility behavior.10

Socioeconomic Achievements of Men and Women

Throughout the project one of our principal efforts has been to develop models of social and psychological influences on educational and occupational attainments of men and women that incorporate our best estimates of errors in parental status variables, social influences, educational and occupational aspirations and attainments. We began our efforts by developing a model for men that incorporates some 26 measured variables into a recursive system of 14 unobservable (latent) constructs; the functioning of 9 of the latter variables is further simplified by postulating 3 other unobservable variables (Hauser, Tsai and Sewell 1983). Briefly, the model specifies that social origins and ability affect post-secondary schooling and occupational careers by way of aspirations and social influences in late adolescence. The analysis asks whether the Wisconsin data are consistent with the modified causal chain hypothesis proposed in the original formulation of the model (Sewell, Haller, and Portes 1969), rather than with models that incorporate many more lagged effects. The causal chain hypothesis receives far greater support than in previous analyses of the data that have not taken account of survey response error (and other stochastic components of latent variables in the model); that is, the lag-1 effects postulated in the model are far stronger than has been found in the past, and few delayed effects are present. For example, the model accounts for 69 percent of the variance in post-secondary schooling, for 73 percent of the variance in the status of first jobs, and for 69 percent of the variance in occupational status at age 36. The model identifies random response errors and correlations among responses obtained on the same occasion, from the same person, or using the same method. The model also allows analysis of the contamination of retrospective reports of social influences and aspirations by intervening events. Thus, the analysis provides new evidence about the stratification process, about the validity of retrospective and contemporaneous reports of status variables, and about the social psychology of retrospection.

This model has also been estimated for women in our sample in order to compare the educational, occupational and economic achievements of men and women (Tsai 1983; also, see Sewell, Hauser, and Wolf 1980). We find that, although women have gained parity in educational attainment, their labor force activities and outcomes are still restricted. Whatever occupational equality may exist at any one stage of the life cycle, women have fewer opportunities for gains in occupational status over the life course. Women obtain smaller returns on their earlier occupational achievement than men do. Whereas women are forced to rely more on academic performance and formal education for occupational placement, men increase their occupational status over the life cycle mainly as a result of their earlier occupational experiences. Moreover, parents transmit direct occupational and economic advantages across generations to their sons, but not to their daughters. On the other hand, women receive larger earnings returns to educational attainment and occupational status than do their male counterparts. The comparisons also indicate that men's earnings are primarily determined by their occupational status, whereas women's earnings are primarily deter-
mined by the amount of labor supplied to the market. Finally, marriage and childbearing have positive effects on men's earnings, but negative effects on women's earnings. However, the negative effects of marriage and childbearing for women disappear when labor force participation is controlled.

Effects of Family Structure and Sibling Resemblance
We have examined the effects of birth order and size of sibship on educational attainment for the full sibships of our primary respondents (Hauser and Sewell 1985). We have undertaken this analysis because of the recent revival of interest in birth order effects resulting from theories proposed by Zajonc and Markus (1975) and by Lindert (1977). In our study of the 30,000 men and women in the full sibships of our 9,000 primary respondents we find no effects of birth order on educational attainment when size of sibship and other relevant variables are controlled, whether we look at selection into the sample of high school graduates, post-secondary educational attainments of those graduates, or educational attainments within full sibships. Educational attainment appears to increase with birth order when family size is controlled but this happens when secular increases in schooling have occurred within as well as across families. Thus, when we control birth year and parental education, there is no significant association between birth order and educational attainment: there are no linear or non-linear effects, there are no effects of being first or last born, and there are no statistically significant or patterned differences among ordinal positions. Thus, there is no need to invoke any of the more complex theories of child development or intra-familial resource allocation to explain the effects of birth order on educational attainment because there is nothing to explain. Retherford and Sewell (1991) have carried out a comprehensive test of the confluence model using data on the mental ability of WLS primary respondents and siblings, and there, too, the findings have been clear and negative.

We have studied sibling resemblance in education, occupational status, and earnings and in age at marriage and fertility (Claridge 1983). We find little resemblance between in fertility between sisters, but there is a great deal of family resemblance in socioeconomic achievement and its antecedents. For example, we estimate that family origins are associated with 49 percent of the variance in measured ability, 46 percent of the variance in educational attainment, 41 percent of the variance in the status of first jobs, 38 percent of the variance in status of current jobs (in 1975), and 27 percent of the variance in earnings. Much of this research has involved the development of structural equation models of sibling resemblance in educational and occupational status (Hauser 1984; Hauser and Mossel 1985; Hauser and Mossel 1987; Hauser 1988). In this work multiple measurements of educational attainment and occupational status for male high school students and their brothers are used to develop and interpret skeletal models of the regression of occupational status on schooling that correct for response variability and incorporate a family variance component structure. These analyses have provided a methodological template for the specification of more complete models of stratification (Hauser and Sewell 1986), and we are very excited about the prospect of extending them to cover the later achievements of WLS respondents and siblings. We have not yet exhausted the possibilities for analyses of sibling resemblance in the existing WLS data, and we are continuing to work on several topics: inter-sibling influence on educational attainment (Lee 1989); the factorial complexity of schooling; sibling resemblance in social participation; and the social psychology of adolescent status attainment. Much of our previous analytic effort has been spent in developing models and methods for these analyses. With the combination of the sibling pair design and multiple measurements obtained from self- and proxy reports by siblings, we believe that it will be possible to make dramatic progress in modeling effects of family background, of individual differences in achievement, and of cross-sibling effects on achievement (Hauser and Wong 1989). We believe that similar models and methods will also help to elucidate social influences on the broader array of outcomes that will be measured in the 1992 survey, especially those pertaining to physical and mental health.

Mental and Physical Health
Psychological well-being will be assessed with a multidimensional formulation of positive functioning based on the integration of clinical, mental health, and life-span developmental theories (Ryff 1989a). The points of convergence in these theories constitute six key dimensions of well-being (autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, self-acceptance), which have been operationalized with structured self-report scales (Ryff 1989b). Preliminary research indicates that the scales have acceptable psychometric properties, and that certain of them, particularly positive relations with others, personal growth, autonomy, and purpose in life, account for additional and independent variance beyond that covered by earlier measures of well-being (e.g., life satisfaction, happiness, self-esteem).

In addition to these instruments, the multifactorial assessment of well-being will include global, single-item indicators as employed in prior survey research (Veroff, Douvan, and Kulk 1981), measures of psychological distress (i.e., depression), and physical health status. These instruments will enable comparisons with other data sets (e.g., ISR Surveys) as well as afford more precise evaluation of the impact of the attainment process on multiple aspects of mental and physical health.
Depression will be measured by the Center for Epidemiological Studies' Depression Scale (CES-D) (Radloff 1977). Physical health will be assessed with the OARS (Duke University 1978) checklist of illness, measures of height and weight, and items regarding subjective health evaluations and perceived changes in health since age 40. Additional items, as developed by the MacArthur Midlife Research Network, have been included to assess activities and time devoted to health maintenance.

The self-evaluation maintenance (SEM) model (Tesser 1980) predicts the conditions under which people will react with either jealousy or pride to the success of comparison others. Specifically, closeness/likeness (e.g., in age, sex) is hypothesized to moderate the effect of relative performance. Thus, the interaction effect predicts that if the sib performs better than the self and the sib is more like the self, there will be greater friction. Using this general model, we can examine the implications of social comparisons in the family for subjective well-being as well as for patterns of support and assistance between family members. We will also examine how sibs cope with discrepancies in their individual achievements. Within-family achievement differentials are hypothesized to be a source of psychological discomfort. To alleviate distress individuals may reconcile their achievements relative to their sibs by discounting the success of comparison of others and relinquishing responsibility for their own shortcomings, e.g., denying responsibility for failure: "I've had little control over the things that happen to me."

Just as people compare their attainments to those of others, they also compare their present selves with their past attainments. Our objective here is to first predict people's present achievements (occupational status, earnings) using the Wisconsin Model. Our interest lies in the effects of achievements, net of individual endowments, education, and achievement in the early career. First, we are concerned with the implications of these differing relative locations for mental and physical well-being in midlife. For example, individuals who have gone beyond their original resources are predicted to show positive self-evaluations (self-acceptance), a sense of effectiveness (autonomy, environmental mastery), and personal progress (purpose in life, personal growth). Second, we are concerned with respondents' aspirations for the future and for their offspring as a function of where they are relative to where they began. Of special interest is whether the success of children helps to mitigate the adverse psychological effects of underachievement among midlife adults. Finally, we will examine how people revise their past as a function of what they have or have not attained. Among the hypotheses derived from control theory is that people rewrite the past. For example, they may look back to an earlier period and recall having lower aspirations than they actually reported at the time. They can thus exaggerate change when in fact little change has occurred (Ross 1989).

Social and Economic Exchanges
In work on household economics with the National Survey of Families and Households, we have been studying the determinants of inter-household transfers (ITs) by focusing on the respective roles of family background, life-course events, and government transfer income opportunities. Our analysis has established that ITs may be particularly important for certain types of households, and especially after age 45. Beyond that age, most NSFH respondents give more in gifts and loans than they receive on average. However a substantial minority reported receiving much more than they gave during the NSFH's 1982-86 recall period. For the 45-59 year subgroup, 12 percent received gifts averaging nearly $9,000, with 7 percent reporting loans that averaged about $10,000. After age 60, the average of gifts and loans received are substantially less—roughly half of their pre-retirement levels. Although the percentage receiving bequests is about 2 percent for all persons over 45 the average amounts of these ITs are large—at about $20,000 for the 45-59 group, and $50,000 for older respondents. Hence although mature adults continue to help their own children via gifts and loans, those who receive help from their relatives (primarily parents) get substantial support from them, and a few can expect to receive very large bequests.

As in our work with the NSFH, we plan to use WLS histories on demographics, earnings, and other experiences to construct variables that describe whether and how recently the respondents had experienced events that would increase (or decrease) their needs for gifts and loans. Events such as the onset of a severe illness influence the timing of these transfers. Given that a transfer occurs, family background, respondent's earnings, and government income opportunities determine how much help gets provided.

For younger persons ITs seem to be associated with major life-course events such as births and marriages that create need for help with basic living expenses. However after age 45 NSFH respondents tend to report that gifts and loans they received were more often for home-buying and other investments, i.e., intended to help them accumulate wealth. Accounting for wealth transfers and their potential influence on well-being and the rigidity of the class structure requires a more comprehensive model that links prior transfers, such as those to fund educational achievement, to current transfer behavior. To study that process we intend to adapt the Wisconsin status attainment model, by elaborating it to include ITs as an influence on status achievements. An NSFH result that motivates our interest in ITs as a potentially impor-
tant intervening variable is that the net effect of respondent’s education on gifts and loans received is highly positive in models that control for father’s education and current earnings. Education may be tapping otherwise unmeasured influences of family wealth operating through educational achievement. However families that provide help to educate their children may continue to support them throughout the life-cycle, for which they presumably get better support in their old age—in which case educational differences tap differences in preferences, not wealth. The family background measures in the WLS are more complete for the purpose of analyzing IT effects than in any other data set. Specifically the parent’s income data from Wisconsin tax records will control much better for initial differences in ability to provide transfers.

Finally, we plan to analyze whether and to what extent both receiving and giving ITs and care-giving assistance influence WLS respondent’s psychological well-being. Douthitt and MacDonald (1990) have been using the Wisconsin Basic Needs survey to study the relative contribution of life-cycle variables and alternative measures of financial status to variation in the Andrews-Withey Delighted-Terrible scale on subjective well-being. As part of that work for NIMH, they have been able to separate the effects of wealth and measures of net worth from current earnings. A WLS follow-up that included a match with Social Security earnings would permit better analyses of the financial sources of variation in global satisfaction measures. Additionally placing perceived well-being as the ultimate dependent variable in a model that includes family background, current economic statuses, and measures of inter-family transfers would yield information about the relative importance of these transfers, as gauged by measures of satisfaction and not merely in dollar terms. In particular, we note that although economists have been very active in modeling the determinants of family assistance, they have not been very explicit about the importance of that assistance—either in economic terms, or as otherwise evaluated by the recipients themselves.

**Design and Methods**

The study is based on a telephone interview and self-administered mail-out, mail-back questionnaire of WLS primary respondents and their siblings. It will build on information about the life course previously obtained in surveys in 1957, 1964, 1975, and 1977 and from various public records. This section describes the means by which new survey information is being collected, integrated with the existing data (excepting confidential Social Security records), subjected to preliminary analyses, and made available to other cooperating researchers as core information on which additional data collection efforts and analyses may be based.

The WLS sample is large and heterogeneous, and it is broadly representative of white American men and women who have completed at least a high school education. The sample is mainly of German, English, Irish, Scandinavian, Polish, or Czech ancestry. Some strata of American society are not well represented in the WLS. Everyone in the primary sample graduated from high school; about 7 percent of their siblings did not graduate from high school. We have estimated that about 75 percent of Wisconsin youth graduated from high schools in the late 1950s. Minorities are not well represented; there are only a handful of African American, Hispanic, or Asian persons in the sample. The WLS sample is sometimes criticized for over-representing persons of farm origins. That is not correct. About 19 percent of the WLS sample is of farm origin, and that is consistent with national estimates of persons of farm origin in cohorts born in the late 1930s. In 1964 and in 1975, about two-thirds of the sample lived in Wisconsin, and about one third lived elsewhere in the U.S. or abroad.11

There has been very little attrition in the course of the WLS. Response rates, relative to the full, initial cohort sample of 10,317, were 86.5 percent in 1964 and 88.6 percent in 1975. (That is, in the 1975 follow-up we did not drop individuals for whom no response had been obtained in 1964.) In the current round of the study, we originally planned to include only the 9138 individuals who participated in the 1975 survey and a surviving sibling (if any) of those individuals. In addition to individuals who died by 1975, this excluded about 3 percent of the original sample who were dropped from the 1975 survey because they could not be found, about 6 percent who were dropped from the sample because they could not be interviewed by telephone (because of illness, institutionalization, or residence outside the U.S., or because they could not be reached by telephone), and another 4 percent of the original sample who refused to participate in the 1975 study.

Before the tracing began (in July 1991), we knew that we would achieve substantial success in tracing the 1975 respondents, for we had found 92 percent of a pilot sample of 184 respondents in the 1975 survey. For the production tracing operation, we divided the sample into three broad strata: 1975 respondents for whom no brother or sister had been drawn into the 1977 sibling survey (about 6500 persons);12 1975 respondents for whom a brother or sister had been drawn into the 1977 sibling survey (about 2500 persons); and 1975 non-respondents who were not known to have died (about 1000 persons). Each of these groups was divided into 10 stratified random replicates. The main lines of stratification are the sex of the respondent and his or her selected sibling, and the educational attainments of the respondent and sibling. We carried out production tracing one
Our tracing efforts are carried out almost entirely by telephone. We begin with a direct call to the primary respondent or selected sibling at the last known telephone number, and we continue with a call to the parents' last known number. Those methods yield sufficient information for about half the cases. We find the remaining cases using a variety of methods, based on previously known addresses, siblings' and children's names, high schools or colleges attended, and places of employment. Two key tools have been a commercial credit union database (in which we have no access to financial information) and a national database of names, addresses, and telephone numbers on CD ROM. We count a case as completed only when we have confirmed names, addresses, and telephone numbers (or lack thereof) for both members of a sibling pair with a responsible adult in their family.

Given the success of the main tracing effort, we decided to carry out a pilot effort to find persons who did not respond in 1975 and were not known to be dead. Using our standard methods we were able to locate 86 percent of a random pilot sample of 99 persons, and — after considering the need to collect additional background material — we decided to include 1975 non-respondents in the new follow-up.

Study Design
The WLS cohort of men and women, born about 1939, precedes by about a decade the bulk of the baby boom generation that continues to tax social institutions and resources at each stage of life. For this reason, the study can provide early indications of trends and problems that will become important as the larger group passes through its fifties. This adds to the value of the study in obtaining basic information about the life course as such, independent of the cohort's vanguard position with respect to the baby boom. In addition, the WLS is also the first of the large, longitudinal studies of American adolescents, and it thus provides our first large-scale opportunity to study the life course from late adolescence through the mid-50s in the context of a complete record of ability, aspiration, and achievement.

Past waves of the study have provided multiple, often overlapping measures of factors affecting life-course aspirations and outcomes. In addition to the fundamental advantage of obtaining true longitudinal measures for causal modeling, multiple measures have been valuable in estimating the effects of measurement error on the parameters of causal models of aspiration and attainment. Of all the multiple measurements, however, the most fruitful have perhaps been the parallel questions asked of core respondents and their siblings. A recent series of papers, described above, has shown the power of this design for discovering the effects of unmeasured factors which operate within families to influence a variety of outcomes in later life (Hauser and Mossel 1985, Hauser and Sewell 1986). Unfortunately, as the possibilities of this feature of the study design have come to seem very promising, the smaller size of the sibling sample (about 2000) compared with the core sample (9,138), has become a limiting factor. Some analyses cannot be done with the low statistical power available at this sample size, for example, when we work with subsamples of sibling pairs defined by the sex of the primary respondent and his/her brother/sister (Lee 1989). For this reason, and because of the substantive importance of investigating family effects, we proposed that a randomly designated sibling of every primary respondent, an additional 5500 persons, be interviewed in this round of the study; at this writing, we expect to have enough support to interview about 4000 of these brothers or sisters, so we will exclude some of the replicate samples from this part of the study.

It is important to note that the existing sample of 2000 siblings was augmented to include all twins of the core sample members, whether or not they had been drawn in the 10,000 original cases of the high school sample. There are 116 distinct pairs of twins, a sizeable number for a sample from a general population, followed for a long period of time.

Timing of Activities
Previous experience with the WLS provided a sound basis for planning the sequence of our activities. The past year was spent primarily on sample tracing, instrument development and pretesting, and the creation of selected abstracts of data from the project files or from the 1975/1977 questionnaires that are being used directly in the 1992 interviews. For example, aside from identifying information, the telephone interviews use prior data on marital status, job in 1975, children, and siblings; we ask the respondent about his relationship with a best high school friend only in the 20 percent of cases where each member of a dyad in the sample named the other as among his or her best high school friends. There are three instruments: the core respondents' interview schedule; the siblings' interview schedule (possibly with some modification for siblings who were not previously
interviewed); and the mail-back questionnaire to be sent both to core respondents and siblings. The first two questionnaires will be very similar. The instruments have been developed and pretested thoroughly with the help of persons in the class of 1957 who are not in the WLS sample.

This year will see the collection of the data, by both telephone and mail, together with additional tracing activities for previously-located respondents who cannot be relocated at the time of the survey. Data will be merged and loaded into a preliminary file, and cleaning operations will begin. As soon as the data are clean, the preliminary files will be made available to interested researchers outside the group. We expect to prepare these files for replicate subsamples on a flow basis, so some data will become available before the fieldwork is complete.

In the third year extensive data merging and variable construction will lead to preliminary data analyses and publications. These early efforts will probably be straightforward exploitations of the new data, and will extend the time horizons in standard sociological and social psychological models of the life course. Also during this year, plans and proposals will be formulated for additional analyses of the data and for the record linkages that will be possible with them.

The WLS Data
The planned research will make use of detailed information already obtained for earlier periods of the life course. Previously collected data span more than 3600 columns of coded items per case, and they cannot be described in detail here. An overview of the existing data may help indicate the potential usefulness of the planned new survey data and linkages.15

In 1962, William H. Sewell obtained data from a 1957 survey of Wisconsin high school seniors in public, private, and parochial schools. A random sample of 10,317 cases (approximately one-third of the seniors), was selected for further study. Information on the measured mental ability of each student was added to the cards from the files of the Wisconsin State Testing Service, which at that time conducted a testing program covering all eleventh graders in the State. A number of indexes based on information from the survey were developed and added to the cards for each student, including the socioeconomic status of the student's family, the student's attitudes toward higher education, educational and occupational plans, and perceived influence of significant others on educational plans.

Relevant measures of school, neighborhood, and community contexts — for example, the socioeconomic composition of each senior class, the percentage of its members who planned on going to college, the size of the school, the size and degree of urbanization of the community of residence, and the distance of the student's place of residence from the nearest public or private college or university — were constructed from secondary sources.

In the spring and summer of 1964, seven years after the students had graduated from high school, we undertook a follow-up study of the original sample. Using a questionnaire on a double postal card, information was obtained from parents on the post-high school education, current occupation, military service, marital status, and present residence of over 87 percent of the sample (Sewell and Shah 1967). With the cooperation of the Wisconsin Department of Revenue (and following their strict arrangements to guarantee the privacy of individual records), information on the parents' occupations and income was obtained from their 1957 to 1960 state income tax returns. Still later, we obtained information on earnings for the males in our sample from the Social Security Administration for each year of covered employment from 1957 to 1967. This phase of the project required an elaborate linkage procedure to protect individual identity. The earnings record was later extended to cover the period from 1957 to 1971. Our data were further enriched by addition from several published sources of information on the characteristics of secondary and post-secondary schools, colleges, and universities.

During 1975, we carried out 1 hour telephone interviews with the sample and obtained the following information from our sample: (1) composition of family of origin: age, sex, and education of each sibling, the occupation and address of a randomly selected sibling, and the parents' ethnic and religious background; (2) the education of the respondent: content, timing, and location of all post-secondary schooling, including vocational, collegiate, and military schooling; (3) labor force experience: dates and types of military service, first civilian job, occupation in 1970, current (1975) job, longest job in 1974, earnings in 1974, weeks and hours worked, location, size, and type of work organization, work satisfaction, work authority, occupational aspirations, labor force participation and jobs held before marriage and in each birth interval (women only); (4) characteristics of family of procreation: marital status, marital history, a roster of children by age and sex, and educational and occupational aspirations for a randomly selected child; spouse's work status, education, occupation, and 1974 earnings; (5) selected retrospective information: aspirations while in high school and names of best high school friends; (6) social participation: membership in organizations, church attendance, visiting behavior, voting. We obtained similar information from interviews with 2000 randomly selected siblings (including all twins) during 1977, and at that time we also searched the records of the State Testing Service for

Spring/Summer 1992
Data Collection

Although no effort was made to obtain an agreement to be interviewed as part of the 1989 trial study, the 1975 survey obtained responses from 88.6 percent of the primary respondents, and the 1977 survey obtained responses from 87.4 of the randomly selected siblings. The project has attempted to cultivate the good will of the sample, through reports made to the respondents and by other means, and we expect that excellent response rates will again be obtained. At this writing, about 700 interviews have been completed in the first two random replicates of the main sample, and these reflect about a 95 percent completion rate among all direct telephone contacts with respondents. Within the first random replicate, the overall response rate is already more than 80 percent.

Survey Operations

The questionnaire will be administered in two parts. Core items dealing with social and demographic characteristics and changes in them, self-assessments of health and well-being, social participation, and relationships with parents, siblings, and children, along with future aspirations and plans, are obtained in the telephone interview, which is being conducted by the University of Wisconsin Letters and Sciences Survey Center (LSSC). Items were selected for the telephone interview if they are administrative required many logical branches or if the items were not grouped with a long list of similar questions. The interview may be somewhat shorter, perhaps 45 minutes, for siblings who participated in the 1977 survey. Interviewers are using computer-assisted (CATI) techniques, with which the lab has long experience. The project staff provides initial telephone numbers to LSSC from its separate tracing activity, and is standing by to do additional tracing when numbers prove to be out of date. Other information essential to the telephone interview, such as rosters of children’s and siblings' names needed for information updates, have been transcribed from the original 1975 questionnaires and entered in the computerized interview schedule database. Responses to occupation and industry questions, which are especially difficult to code, are routed from the field to our occupation coding section, and cases with incomplete responses are returned to the field within a day or two for a callback.

One useful feature of the CATI interview is the ability to introduce alternate forms or to sample selected questions at different rates in different internal replicates. For example, we are using two different series of questions about job authority, each administered to half the sample; we are asking a lengthy set of questions about depression and alcohol use of 80 percent of the sample; and we are asking about the current income of surviving parents in half the sample. We may alter sampling rates of these and other questions as the fieldwork proceeds.

Because the telephone interview should not be too long, some of the social psychological, health, occupational, and social exchange data are being obtained with a mailed, self-administered questionnaire. Mail items tend to be groups of closely related questions with few logical contingencies and similar closed-ended response alternatives. The mailed questionnaire requires about 30 to 45 minutes to complete. LSSC is providing two remailings to encourage respondents to mail back their questionnaires. However, a subset of the items in the psychological scales is administered in the initial telephone interview, to avoid a total loss of information from those not returning the mail questionnaire. Appropriate statistical techniques will allow the resulting partial information to be included in structural models with measurement error, correcting for biases that would otherwise be intractable (Allison 1987, Allison and Hauser 1991). After two pretests of preliminary mail questionnaires, we carried out a final pilot test of the mail instrument with three waves of mailing, and we obtained an 80 percent response rate.

As explained above, we hope that respondents will grant us a limited, written waiver for the use of their Social Security numbers (SSN’s) to obtain Social Security earnings data. This will permit us to use our existing files of Social Security data directly, and, more important, it will permit us to build earnings histories of women, to complete the earnings histories of men in the WLS, and to obtain additional data from Social Security records on disability, dependency, and death. Aside from the administrative requirement to have written waivers for access to these data, the SSN will also be important in linking the WLS to the National Death Index in future studies of differential mortality. We have SSN’s for almost all of the males but for none of the females in the WLS; we have no waivers at all. We want to obtain waivers and additional SSN’s without losing the good will of the sample.

Our tentative plan for obtaining waivers is as follows: During the telephone interview, we ask the respondent to give us his or her social security number (SSN). If the response to this request is negative, the matter will be dropped; if it is positive, as it is in some 92 percent of the interviews completed thus far, we will mail a waiver form after completion of the mail interview. The mailed waiver itself will be accompanied by a note inviting the respondent to call the principal investigator directly with any questions. We had originally planned to obtain waivers before completing the mail interviews, but delays in reaching an agreement with the Social Security Administration have precluded this design.
The replicate samples will be introduced sequentially into the interviewing, mail survey, and waiver processes, just as in the tracing operation. Aside from the advantages already mentioned, a smooth work flow and feedback on response rates, this design makes it possible to terminate data collection prematurely if costs run above budget; that is, it will be possible to reduce costs by lowering the size of the final sample without jeopardizing the quality of the data or permitting nonresponse rates to rise to an unacceptable level. The two thousand matched sibling pairs for whom we already have sibling interviews (from 1977) are in pre-existing subsets of the WLS; they will be introduced into the field operations near the beginning of the process, but not at its very beginning. That is, we want to be sure that everything is working smoothly before we begin to work on these key segments of the sample, but we do not want to wait so long that there is any chance of our terminating the field operations before their data have been collected. A similar internal sampling procedure was used successfully in the 1975 follow-up survey.

1. The research described herein is supported by grants from the National Institute on Aging and the National Science Foundation and by the Graduate School of the University of Wisconsin-Madison. Preparation of this paper was supported in part by the Spencer Foundation, the William F. Vilas Trust, and the Kenneth and Carolyn Brody Foundation and by a training grant from the National Institute on Aging to the Center for Demography and Ecology at the University of Wisconsin-Madison. The opinions expressed herein are those of the authors. Please address all correspondence to Robert M. Hauser, Department of Sociology, The University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706.


3. We began this round of study with the intention of following only those individuals who had been interviewed in 1975. However, we found it possible to locate previous non-respondents, as well, and we now plan to follow and interview all surviving members of the original sample.

4. One might think of mental ability, educational attainment, or occupational prestige as general effects, whereas aptitude, personal contact with an entrepreneur, or training in cosmetology are local effects.

5. These data (with identifiers removed) have been placed in the public domain through the Data and Program Library Service of the University of Wisconsin-Madison. One exception is Social Security Earnings histories of men in the sample from 1957 through 1971, which were obtained under conditions which preclude their distribution (or the direct access of the investigators to the data files in which they are contained). Two other exceptions are files of detailed characteristics of colleges attended and of the employers of the primary respondents in 1975; these are not confidential, but we maintain them separately from the master file.

6. The WLS data have been used in 4 research monographs, 23 doctoral theses, 11 masters theses, and more than 100 research articles or chapters in books. Sewell and Hauser (1992) review the study from the early 1960's to the present.

7. In the first 400 completed interviews, the rates of parental survivorship far exceeded our estimates: 55 percent of respondents had a living mother, and 26 percent had a living father. These cases represent the first 62 percent of persons to respond within a stratified random subsample of 650 primary respondents.

8. A pilot effort to relocate members of the Project Talent sample, carried out in parallel with our initial feasibility tests, provided unsatisfactorily low coverage.

9. The Wisconsin Longitudinal Study was supported continuously by the National Institute of Mental Health (MH-6275) from 1962 through 1982. During that period, the WLS also obtained support from the Social Security Administration (Social and Rehabilitation Service Grant No. 314) for linking and analyzing earnings histories and from the Spencer Foundation for the 1977 survey of siblings. From 1980 to 1986 the project was supported by NSF grants for studies of sibling resemblance (SES 80-10640) and for the documentation of machine-readable data (SES 83-19879). The WLS had no federal support from 1986 to 1991, and we have continued to work on analyses of family effects on achievement with support from the Guggenheim Foundation, the Volkswagen Foundation, the Graduate School of the UW-Madison, the Brody Foundation, the Spencer Foundation, and the use of core facilities of the Center for Demography and Ecology at the UW-Madison, which are supported by grants from the National Institute of Child Health and Human Development (HD-5876) and the William and Flora Hewlett Foundation.

10. This text covers only a few of the issues in recent WLS research. For a full review see Sewell and Hauser (1992).

11. The 1991-92 respondent tracing activity show a similar distribution of respondents between Wisconsin and other locations.

12. We had selected a brother or sister of these persons during the 1975 survey, but we could not afford to
interview them at that time.

13. These individuals were designated in the course of the 1975 interview with the primary respondents, and at that time their full name, address, age, sex, educational attainment, occupation, and industry were ascertained, along with the name of the last Wisconsin high school they were known to have attended. The last piece of information is helpful in finding mental test scores. Thus, while the records for these individuals will lack the self-reported information obtained in the 1977 sibling interviews, there is already some baseline information about them in the WLS files. Funding for this phase of the study not been obtained.

14. Copies of the mail questionnaire and a list of questions in the telephone interview are currently available from the authors. At this time, there is no complete written text for the telephone interview, other than the script for the CATI program used in the survey (CASES).

15. With the exception of identifiable or confidential material, these data are now available from the Data and Program Library Service of the University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706. We expect to release the new edition of the data through the Inter-university Consortium for Political and Social Research.

REFERENCES


36 IASSIST Quarterly


IASSIST Constitution
As Amended December 15, 1992

ARTICLE I - NAME
The name of this organization shall be the INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICES AND TECHNOLOGY/ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES, hereafter referred to as "IASSIST".

ARTICLE II - HEADQUARTERS
The official headquarters of IASSIST will be located with the Treasurer.

ARTICLE III - OBJECTIVES
All activities of IASSIST will be based upon the following objectives:

3.1 To encourage and support the establishment of local and national information centers for social science machine-readable data.

3.2 To foster international exchange and dissemination of information regarding substantive and technical developments related to social science machine-readable data.

3.3 To coordinate international programs, projects, and general efforts that provide a forum for discussion of issues relating to social science machine-readable data.

3.4 To promote the development of standards for social science machine-readable data.

3.5 To encourage educational experiences for personnel engaged in work related to these objectives.

ARTICLE IV - ACTIVITIES
To accomplish the objectives of IASSIST, some or all of the following activities may be conducted with the approval of the Administrative Committee on a national or regional basis and the submission of an appropriate report:

4.1 COMMITTEES AND GROUPS

Committees may be established and groups of members organized to undertake specific tasks, to find solutions to specific problems, to develop and compile relevant material for specific projects, and to disseminate information on specific subjects.

4.2 CONFERENCES, WORKSHOPS, SEMINARS, TRAINING SESSIONS
Members may convene organized efforts on any subject consistent with IASSIST objectives.

4.3 PUBLICATIONS

A Newsletter will be published and regularly circulated to all members, as well as to others wishing to subscribe. Other kinds of publications may be produced on occasions.

4.4 COOPERATION WITH OTHER ORGANIZATIONS

Efforts will be made to cooperate with other organizations in joint projects and activities when these are consistent with IASSIST objectives.

4.5 OTHER

Other activities that advance the objectives of IASSIST may be undertaken from time to time.

ARTICLE V - MEMBERSHIP

5.1 The membership shall consist of regular and student members, and shall be open to such persons as are interested in supporting the objectives of IASSIST.

5.2 Membership in IASSIST shall include a subscription to the Newsletter.

5.3 Resignations of any members shall become effective immediately upon receipt by the Treasurer of IASSIST. Resignation shall imply forfeiture of the annual membership fee.

ARTICLE VI - FINANCES

6.1 The fiscal year of IASSIST shall begin 1 January and end 31 December.

6.2 Membership fees for regular and student members shall be paid annually to the Treasurer by 1 March of each fiscal year.

6.3 The rate of membership fees may be changed by a two-thirds vote of the members on a mail ballot or during the Business Meeting of the General Assembly. Mail ballots will be undertaken between October and December of any calendar year. The results of such ballots or votes will go into effect on 1 March of the following year. In the event of a vote during the Business Meeting of the General Assembly, the membership will be informed prior to the Business Meeting and proxy ballots will be made available.

ARTICLE VII - GOVERNANCE

7.1 GENERAL ASSEMBLY

IASSIST shall consist of a General Assembly composed of all regular and student members. The General Assembly will be organized by geographic regions. The establishment of a region must be approved by the Administrative Committee.
7.2 FUNCTIONS OF THE GENERAL ASSEMBLY

The General Assembly will establish general policies for IASSIST and elect the members of the Administrative Committee, as well as the officers of the Association. Each region will, in addition, elect its own administrative officer who will be known as the Regional Secretary.

7.3 ADMINISTRATIVE COMMITTEE

The Administrative Committee will be the executive body of IASSIST, and shall be composed of at least 10 members elected by the General Assembly from its membership. The composition of the Administrative Committee will reflect the geographic distribution of the members of IASSIST and will be based on the number of members in each geographic region; the Regional Secretaries; the immediate past-President of IASSIST; the President and Vice-President; and the Treasurer, the Editor, the Secretary, and the Archivist, the last three individuals having been appointed by the President with approval of the Administrative Committee. The elected members of the Administrative Committee will serve a four-year term and may serve no more than three consecutive terms. The Regional Secretaries will serve a two-year term and may serve no more than three consecutive terms.

7.4 FUNCTIONS OF THE ADMINISTRATIVE COMMITTEE

The Administrative Committee will implement policies, develop future directions, and coordinate activities for IASSIST. The Administrative Committee will organize the General Assembly into geographic regions, determine the number of Administrative Committee members from each geographic region, and call meetings of the General Assembly at least once every year. The Administrative Committee will also establish Committees and Groups as required.

7.5 OFFICERS OF THE ASSOCIATION

The Nominations Committee will propose candidates for the offices of President and Vice-President, to be voted upon by the General Assembly. These officers shall serve a two-year term and may serve no more than three consecutive terms.

7.6 ROLE OF THE OFFICERS

The officers of IASSIST will be responsible for the conduct of business of the ASSOCIATION between meetings of the Administrative Committee.

7.7 EXECUTIVE COMMITTEE

The Executive Committee will consist of the Officers, plus other members of the Administrative Committee as required and designated by the Officers.

ARTICLE VIII - MEETINGS

8.1 The annual meeting of the General Assembly shall be held at a time and place chosen by the Administrative Committee.

8.2 Special meetings of the General Assembly may be called by the Administrative Committee.

8.3 The Secretary shall give notice to the members as to the time and place of the annual meeting or special meeting not less than two months prior to the scheduled meeting.

8.4 A quorum shall consist of 40 members.
ARTICLE IX - ELECTIONS

9.1 A Nominations and Elections Committee will be appointed by the Administrative Committee.

9.2 The Nominations and Elections Committee shall conduct an election in each geographic region for officers of IASSIST, members of the Administrative Committee, and the Regional Secretaries. Members within each designated geographic region shall only be entitled to nominate and vote for the Regional Secretary in their home region. However, all members will be entitled to nominate and vote for the officers of IASSIST and the other members of the Administrative Committee.

In the event that competitive circumstances do not exist for a Regional Secretary may be appointed by the Administrative Committee.

9.3 A public call for nominations will be sent out by the Nominations and Elections Committee. Voting will be conducted by mail ballot. Elections will be held every two years.

ARTICLE X - AMENDMENTS

The Constitution of IASSIST may be amended by a two-thirds vote of the members on a mail ballot, such ballots to be undertaken between October and December of any calendar year, the results of such ballots to go into effect at the following year’s annual meeting of the General Assembly, provided that:

10.1 notice of the proposed amendments shall have been given in writing to the Standing Committee on Constitutional Review with the written support of at least five (5) members in good standing of the ASSOCIATION; and

10.2 two months’ notice of the proposed amendments is given in writing to all members of the ASSOCIATION prior to the conduct of the mail ballot.

ARTICLE XI - TERMINATION

IASSIST may be dissolved by a majority of the members. All property and funds of IASSIST will be transferred to a branch of UNESCO to be determined by the Administrative Committee.

ARTICLE XII - BY-LAWS

SECTION I DUTIES OF THE PRESIDENT

12.1 The President shall

i. be the principal officer of IASSIST;

ii. provide leadership and guidance in the realization of IASSIST’s objectives;

iii. preside at all meetings of the General Assembly and the Administrative Committee;

iv. be an ex-officio member of all Standing Committees and shall coordinate their activities;

v. represent IASSIST in its dealings with external bodies and agencies, particularly those at the international level; and

vi. report on the state of IASSIST at each annual meeting of the General Assembly.
SECTION 2 DUTIES OF THE VICE-PRESIDENT

12.2 The Vice-President shall:

i. perform the duties and exercise the powers of the President in the absence or disability of the latter;

ii. assist the President in recommending measures to further the objectives of IASSIST when and as often as requested;

iii. be an ex-officio member of all Action and Interest Groups and coordinate their activities, and be responsible for proposing the Coordinators to the Administrative Committee and maintaining regular contact with such Action and Interest Groups throughout the year; and

iv. in the event of the resignation, death, or incapacity of the President, succeed as acting President for the duration of the then President’s term.

SECTION 3 DUTIES OF THE REGIONAL SECRETARIES

12.3 The Regional Secretaries shall:

i. be the primary officers of IASSIST in their respective regions, working closely with the President of IASSIST;

ii. provide leadership and guidance in the realization of IASSIST’s objectives in their respective regions;

iii. represent IASSIST in its dealings with external bodies and agencies, particularly those at the national level;

iv. serve as members of the Standing Committee on Membership;

v. attend all meetings of the General Assembly and the Administrative Committee; and

vi. work closely with the Program Director of the Annual Meeting when the latter is scheduled in their particular region.

SECTION 4 DUTIES OF APPOINTIVE OFFICIALS

12.4.1 The Secretary shall:

i. be appointed by the President of IASSIST with the approval of the Administrative Committee.

ii. attend meetings of the Administrative Committee and meetings of the General Assembly and shall record all facts and minutes of all proceedings in the books kept for that purpose;

iii. be an ex-officio member of the Nominations and Elections Committee to maintain lists of nominees for office and to assist in the preparation and distribution of ballots;

iv. be an ex-officio member of the Standing Committee on Constitutional Review to maintain notices of proposed amendments to the Association’s constitution and to assist in the preparation and distribution of ballots;

v. give notice of all meetings of the General Assembly and of the Administrative Committee or President.
12.4.2 The Treasurer shall:

i. be appointed by the President of IASSIST with the approval of the Administrative Committee.

ii. have the custody of the funds and securities of IASSIST and shall keep full and accurate accounts of receipts and disbursements in books belonging to IASSIST and shall deposit all monies and other valuable effects in the name and to the credit of IASSIST and in such depositories as may be designated by the Administrative Committee from time to time;

iii. disburse the funds of IASSIST as may be ordered by the Administrative Committee;

iv. render to the Administrative Committee at its various meetings, or whenever the members of the Administrative Committee may require it, an account of all his/her transactions as Treasurer and of the financial position of IASSIST;

v. prepare a written report for submission to the General Assembly at its annual meeting;

vi. provide the Standing Committee on Membership with up-to-date mailing lists of all members in good standing in each of the geographic regions;

vii. perform such other duties as may from time to time be determined by the Administrative Committee.

12.4.3 The Editor of the Newsletter shall:

i. be appointed by the President of IASSIST, on the advice of the Standing Committee on Publications and with the consent of the Administrative Committee, for a term of two calendar years which may be renewed;

ii. serve on the Standing Committee on Publications; and

iii. be responsible for the regular preparation, publication, and distribution of IASSIST’s official Newsletter.

12.4.4 The Program Director of the Annual Meeting shall:

i. be appointed by the President of IASSIST with the consent of the Administrative Committee;

ii. set up and organize the next annual meeting following the appointment;

iii. be responsible for keeping the Administrative Committee regularly informed of all preparations; and

iv. work closely with the Regional Secretary in the region in which the annual meeting is to be held.

12.4.5 The Archivist shall:

i. be appointed by the President of IASSIST with the approval of the Administrative Committee;

ii. solicit and obtain records and other documentary materials from former and current officers and from the general membership in order to document the policies, procedures, and transactions of IASSIST;

iii. maintain these materials in an archives or arrange for their orderly transfer to another archives designated by the Administrative Committee; and

iv. take action to promote use of these materials.
SECTION 5 COMMITTEES

12.5.1 The Administrative Committee at the time of the annual meeting of the General Assembly shall appoint and/or confirm Standing Committees and shall appoint and/or confirm Chairpersons of the said Standing Committees.

12.5.2 Standing Committees shall advise the Administrative Committee on matters of policy within their particular sphere, and shall have a Chairperson appointed for a two-year term which may be renewed, two members drawn from the regular membership of IASSIST appointed for a two-year term which may be renewed, one member of the Administrative Committee appointed for a two-year term which may be renewed unless representation from the Administrative Committee is already included in the composition of the Standing Committee in another capacity, and such officers as are designed ex-officio members.

12.5.3 The Standing Committees of IASSIST are the following:

i. CONSTITUTIONAL REVIEW COMMITTEE: responsible for receiving proposals for the enacting, amending, and repealing of the by-laws of IASSIST and for preparing revised articles and by-laws for members’ approval, as well as for undertaking an annual review of the constitution and by-laws and proposing amendments as it deems appropriate.

ii. EDUCATION COMMITTEE: responsible for the development and advancement of professional programs in education and training and for advising the Administration Committee on the criteria for the approval and certification of such programs.

iii. MEMBERSHIP COMMITTEE: responsible for recruiting membership in IASSIST and for recommending alterations in the classes of membership and dues. This Committee’s membership shall include the Regional Secretaries.

iv. NOMINATION AND ELECTIONS COMMITTEE: responsible for receiving nominations for the election of the Administrative Committee, the Regional Secretaries, and the officers of IASSIST, distributing ballots and electoral information according to regulation, tallying the ballots, reporting on the results of the tally, and for recommending alterations in procedures.

v. PUBLICATIONS COMMITTEE: responsible for advising the Administrative Committee on general publications program policy and for reviewing manuscripts submitted for publications. This Committee’s membership shall also include the Editor of the Newsletter.

SECTION 6 ACTION GROUPS

12.6.1 The Administrative Committee, at the time of the annual meeting of the General Assembly, may appoint Action Groups and for every Action Group so appointed a Coordinator shall be named.

12.6.2 A minimum of three (3) members of IASSIST may make application to the Administrative Committee for the establishment of an Action Group at least one month prior to the annual meeting of the General Assembly.

12.6.3 Action Groups shall be expected to undertake specific tasks, to find solutions to specific problems, or to develop and compile relevant materials for specific projects. The mandate or terms of reference of Action Groups shall be clearly defined, including the resources and time required and the specific nature of the output or product.

12.6.4 Action Groups shall report to the Administrative Committee through the Vice-President on matters relating to their particular sphere, and shall have a Coordinator appointed for a one-year term which may be renewed, two or more members of IASSIST appointed for a one-year term which may be renewed, and such officers as are designated ex-officio members.

Spring/Summer 1992 45
SECTION 7 INTEREST GROUPS

12.7.1 The Administrative Committee, at the time of the annual meeting of the General Assembly, may appoint Interest Groups and for every Interest Group so appointed a Coordinator shall be named.

12.7.2 A minimum of five (5) members of IASSIST may make application to the Administrative Committee for the establishment of an Interest Group at least one month prior to the annual meeting of the General Assembly.

12.7.3 Interest Groups shall be expected to disseminate information on specific subjects and to serve as a forum of discussion between as well as during annual meetings.

12.7.4 Interest Groups shall report to the Administrative Committee through the Vice-President on matters relating to their particular sphere, and shall have a Coordinator appointed for a one-year term which may be renewed, four or more members of IASSIST appointed for a one-year term which may be renewed, and such officers as are designated ex-officio members.

SECTION 8 NOMINATIONS AND ELECTIONS PROCEDURES

Any regular member in good standing is eligible to hold office in IASSIST.

12.8.1 The Administrative Committee and the Officers.

i. Every two years, the President, the Vice-President and one-half of the elected members of the Administrative Committee shall be elected from a slate of candidates put forward by the Standing Committee on Nominations and Elections.

ii. During the fall of any election year, any member in good standing may submit in writing to the Nominations and Elections Committee, the names of as many as seven (7) persons for the slate of candidates regardless of the geographic region in which the nominees reside.

iv. The Nominations and Elections Committee will compile a list of nominees which shall be reviewed by the Administrative Committee and will mail ballots to the membership during the fall/winter of any election year.

v. All members in good standing, regardless of the geographic region in which they reside, shall be eligible to vote for a limited number of nominees from each geographic region. The number of nominees from each region will be specified on the ballot, based on each region's percentage of the total membership of IASSIST. Voting will take place over a period of one month during any election year, but in no instance will it extend beyond mid-December.

vi. The results of the election shall be announced by the end of December in every election year. The results shall be published in the first issue of the Newsletter following the election.

vii. Newly elected members of the Administrative Committee and the Officers shall take office after the annual meeting of the General Assembly following the elections.

12.8.2 The Regional Secretaries

i. Every two years, the Regional Secretaries shall be elected from a slate of candidates put forward by the Standing Committee on Nominations and Elections.

ii. During the fall of any election year, any member in good standing in a particular geographic region may submit in writing to the Nominations and Elections Committee, the name of a person for Regional Secretary who must reside in the same geographic region as the nomination.
iii. A nomination must be accompanied by a written statement from the nominee declaring his/her willingness to stand for election; a statement indicating that the nominee has institutional support to undertake the duties; and an outline of the qualifications of the nominee.

iv. The Nominations and Elections committee will compile lists of nominees and mail appropriate ballots to the membership of each geographic region the fall/winter of any election year. v. All members in good standing in each geographic region shall be eligible to vote for the Regional Secretary for that particular geographic region. Voting will take place over a period of one month during any election year but in no instance will it extend beyond mid-December.

vi. The results of the election shall be announced by the end of December in every election year. The results shall be published in the first issue of the Newsletter following the election.

vii. Newly elected Regional Secretaries shall take office after the annual meeting of the General Assembly following the elections.
We mailed out 143 ballots. We had 69 returned for offices and 67 for the proposed amendment to the Constitution. The following persons are elected:

**PRESIDENT**
Chuck Humphrey, University of Alberta

**VICE-PRESIDENT**
Elizabeth Stephenson, UCLA

**ADMINISTRATIVE COMMITTEE**

**U.S.A.**
Carmen Campbell, Bureau of the Census
Jo Ann Dionne, Yale University
Jean Stratford, University of California at Davis

**Europe**
Vigdis Kvalheim, Norwegian Social Science Data

**Canada**
Hilde Colenbrender, University of British Columbia

**REGIONAL SECRETARIES**

**U.S.A.**
Ann Lightfoot Cooper, University of Wisconsin

**Canada**
Wendy Watkins, Carleton University

**Europe**
Peter Burnhill, University of Edinburgh

**Australia**
Roger Jones, Australian National University

---

Vote on the constitutional amendment to separate the office of the Secretary-Archivist into two offices: All votes for the amendment.
The International Association for Social Science Information Service and Technology (IASSIST) will hold its 19th annual conference in conjunction with the International Federation of Data Organisations (IFDO) in Edinburgh over the period 11 - 14 May 1993. This is the first time that the conference has been held in the UK.

The Conference spans the 3 days 12 - 14 May and addresses the concern of IASSIST and IFDO members for managing and sharing computer-readable data during a time of rapid change. This theme highlights the value of openness in sharing data, the richness of diversity among data sources and the standards by which data might be exchanged across disciplinary and national boundaries.

**Preliminary Conference Programme (2nd Revision)**

**Workshop Day Tuesday 11 May**
**Morning**
Concurrent sessions

- What is a data library & how to start one
- UNIX: introduction
- Storage media & handling multimedia collections

**Afternoon**
Concurrent sessions

- Computer-readable data - a challenge for the archivist
- Advanced UNIX: shell scripts, perl, awk, admin & user utilities
- Internet - making contact with networked data resources

**Evening**
Welcome reception

**Day 1 Wednesday 12 May**
**Morning**

**Plenary Session**  
OPENNESS & ACCESS

Concurrent sessions

- Open access to public data
- Bibliographic Control of Computer Files: past & future, an international report
- Data libraries - the new, the reformed & the specialised: delivering data for secondary analysis
Afternoon
Concurrent sessions

Service access to national Census data
Access to electronic records: automation & archives
Views on metadata

Evening
Icebreaker evening

Day 2 Thursday 13 May
Morning
Plenary Session WELCOME DIVERSITY

Concurrent sessions

International comparative research
Diverse uses of electronic records
Diverse media for data exchange and / or service delivery

Roundtable lunches Discussion groups on pre-set topics

Afternoon

Panel discussion Diverse or convergent paths What course for social science data archives & data libraries, including ‘New Directors for Old Archives’

Poster sessions

Evening Conference dinner & Ceilidh

Day 3 Friday 14 May
Morning

Plenary Session COOPERATION THROUGH STANDARDS

Concurrent sessions

Standards for metadata & data documentation: data creation & publication
The electronic book - new standards
Cross-national standards for industrial & occupational classifications

IASSIST Business Lunch

Afternoon

Concurrent sessions

Standards for metadata & data documentation: computer-readable codebooks & delivery to the user
Archival responsibility for research data: towards professional standards
Standards for data exchange & access: words and images

Evening Depart for Highland Weekend
Conference Registration Fees | Pounds Sterling
---|---
Conference & Workshop | 150 *
Conference only | 130 *
Workshop only | 75
One Day Attendance | 75

(* Deduct 25 pounds if you have paid 1993 IASSIST subscription at time of booking.)

Special events

Welcome reception Tuesday 11 May
The Keeper of the Scottish Record Office has kindly invited us to hold our Welcome Reception in HM General Register House on Princes Street. The Presidents of IASSIST and IFDO will welcome delegates, to meet new and intending members and to renew friendships with existing members of both organisations.

Icebreaker Wednesday 12 May

The evening starts with a tour of Edinburgh, highlighting the city's historic and sometimes murky past. We then cross the River Forth to the Queensferry Lodge Hotel in the Kingdom of Fife for a buffet meal. On the return journey we shall have an excellent view of the illuminations on the Forth Bridge. (The cost to non-delegates is 15 pounds)

Conference dinner Thursday 13 May

This will be held in the splendid Assembly Rooms in the Georgian part of Edinburgh, a short walk from the Carlton Highland Hotel. For this we have arranged a whisky tasting followed by a meal, and then a live band, country dancers and traditional Scottish entertainers will lead us into a Ceilidh. (The cost to non-delegates is 20 pounds)

Highland weekend Friday 14 May to Monday 17 May

A post-conference weekend has been organised in the Scottish Highlands, staying at the Isles of Glencoe Hotel at Ballachulish, Argyll. The cost of the weekend is an additional 135 pounds for twin / double rooms, 182 pounds for single room and 53 pounds for a child sharing parent's room. Further information is given in the accompanying leaflet. Early booking is advised as places are strictly limited.

Edinburgh attractions

The ancient and historic Capital of Scotland, Edinburgh is a superb location in which to attend a conference: it is close to both sea and hills and reputed to offer the 'best quality of life in the UK'. In addition to its famous Castle, Edinburgh is renowned for its fine art galleries, Royal Botanical Gardens, Zoo and numerous architectural as well as gastronomic and musical delights. The University of Edinburgh, founded in 1583, has advanced computing and networking facilities; it is a major UK university with an international reputation in the social sciences and the new information sciences.
Transport

There are regular air services from USA, Canada and Europe to Scotland (Glasgow & Edinburgh) and also connecting flights to Edinburgh from London (Gatwick, Heathrow or Stansted). There is a frequent bus service from Glasgow Airport to Edinburgh. The train journey from London to Edinburgh takes under five hours.

Accommodation

The Conference will be held in the Carlton Highland Hotel which is situated near the Royal Mile in the centre of Edinburgh. Rooms have been reserved and special rates negotiated for IASSIST/IFDO’93 delegates at the Carlton Highland, single or twin rooms, and at two other nearby hotels, the Scandic Crown and the Old Waverley, with a special rate for single rooms only.

The allocation of rooms is held until 28th February only and early booking is advisable.

IASSIST/IFDO’93 rates in pounds sterling

<table>
<thead>
<tr>
<th>Hotel</th>
<th>Twin room per night</th>
<th>Single room per night</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlton Highland Hotel</td>
<td>90</td>
<td>68</td>
</tr>
<tr>
<td>North Bridge, Edinburgh EH1 1SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel: +44 (0)31 556 7277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax: +44 (0)31 556 2691</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scandic Crown Hotel</td>
<td>-</td>
<td>68</td>
</tr>
<tr>
<td>80 High Street, Edinburgh EH1 1TH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel: +44 (0)31 557 9797</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax: +44 (0)31 557 9789</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old Waverley Hotel</td>
<td>-</td>
<td>60</td>
</tr>
<tr>
<td>Princes Street, Edinburgh EH2 2BY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel: +44 (0)31 557 4648</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax: +44 (0)31 557 6316</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please note that prices include a full Scottish breakfast and local taxes and that tea and coffee making facilities are available in every room. Bookings should be made directly with the hotel of your choice, quoting the IASSIST/IFDO’93 Conference.

Alternative accommodation

There are many other hotels, guest houses, bed & breakfasts and self-catering establishments located in central Edinburgh. For further information please contact

Edinburgh Marketing Central Reservations Department
3 Princes Street, Edinburgh EH2 2QP
Tel: +44 (0)31 557 9655
Fax: +44 (0)31 557 5118

Conference Personnel

Charles Humphrey, IASSIST President
Paul de Guchteneire, IFDO President
Peter Burnhill, Programme Committee Co-Chairman (IASSIST)
Eric Tanenbaum, Programme Committee Co-Chairman (IFDO)
Alison Bayley, Local Arrangements Committee Chairman
IASSIST/IFDO’93
19th Annual International Conference, Edinburgh, May 11-14 1993
Booking Form (please use one form per delegate & photocopy as necessary)

<table>
<thead>
<tr>
<th>Family name</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First name/initials</th>
<th>Title</th>
<th>Fax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organisation / Institution / Company</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>Postal / zip code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Registration (pounds sterling)**

- **Conference & Workshop** 150.
- **Conference only** 130.
- **Workshop only** 75.
- **One day attendance on Conference** 75.
- **One day attendance on Workshop** 75.
- **Non-delegate ticket(s) for Icebreaker** @ 15.
- **Non-delegate ticket(s) for Conference Dinner** @ 20.
- **Non-delegate membership of University Club** @ 2.75

<table>
<thead>
<tr>
<th>Deduct 25 if you have paid 1993 IASSIST membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

I enclose / shall arrange payment in pounds sterling as follows:

- [ ] Personal cheque payable to IASSIST/IFDO’93 (not acceptable from France)
- [ ] Eurocheques, each not to exceed 100 pounds, payable to IASSIST/IFDO’93
- [ ] Bank Transfer to Bank of Scotland, 32A Chambers Street, Edinburgh EH1 1HU, Scotland, UK
  for credit of IASSIST/IFDO’93 Account 00135889 at Branch 80-02-24, Telephone +44 (0)31 243 5870
- [ ] Visa [ ] Mastercard
  number ____________________________
  expiry date _______________________

<table>
<thead>
<tr>
<th>Billing address (if different from above)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Signature ____________________________ date ____________

To be returned to: Alison Bayley, Data Library, Main Library, George Square, Edinburgh EH8 9LJ, Scotland
The International Association for Social Science Information Services and Technology (IASSIST) is an international association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the IASSIST QUARTERLY. They also benefit from reduced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:
- Regular Membership: $40.00 per calendar year.
- Student Membership: $20.00 per calendar year.
- Institutional Subscriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subscription:
- $70.00 per calendar year (includes one volume of the Quarterly)

Please make checks payable to IASSIST and mail to:
Mr. Marty Pawlock
Treasurer, IASSIST
% 303 GSLIS Building, Social Science Data Archives, University of California, 405 Hilgard Avenue, Los Angeles, CA 90024-1484

I would like to become a member of IASSIST. Please see my choice below:
- $40 Regular Membership
- $20 Student Membership
- $70 Institutional Membership

My primary Interests are:
- Archive Services/Administration
- Data Processing
- Data Management
- Research Applications
- Other (specify)

Name / title
Institutional Affiliation
Mailing Address
City
Country / zip / postal code / phone