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 Piovesan, Research Data Library, W.A.C.-Bennett Library, Simon Fraser University, Burnaby, B.C., V5A 1S6 CANADA (604) 291-4349 E-Mail: USERDLIB@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.UCI.EDU

Title: Newsletter - International Association for Social Science Information Service and Technology
ISSN - United States: 0739-1137 Copyright 1985 by IASSIST. All rights reserved.
Distributed Database Systems

by Raymond Board
Division of Research and Statistics Federal Reserve Board Washington, DC 20551

abstract
A distributed database system is a collection of logically-related databases that are connected by a communications network, together with a software system for managing and accessing the data. A distributed database system is designed so that it appears to the user to be a single, unified database. This paper reviews the advantages and disadvantages of distributed database systems, and discusses issues relevant to their design and use. These issues include concurrency control, distributed query processing and transaction management, disaster recovery, reliability, and methods for distributing data.

Background
Much social science research involves the use of large datasets. These datasets are usually shared among many researchers, each of whom periodically wants to read, update, correct, and add to the data. Managing such datasets presents significant problems, and maintaining them is not an easy task. Frequently, specialized computer software systems, known as database management systems, are used to handle large datasets accessed by multiple users.

A recent trend in computing has been to move away from the traditional mainframe environment toward a distributed computing environment. In a mainframe environment, all users share the same large, powerful computer. In a distributed environment, many computers — usually smaller machines such as personal computers (PCs) — are linked together by a communication network. Such networks can connect users working on different types of machines and in different locations.

This paper discusses an emerging computer technology that is intended to solve many of the problems associated with large datasets. This technology, called database management systems, takes advantage of a number of the benefits offered by distributed computing environments. As this is a relatively new technology, the reader should note that not all of the features of distributed database systems are currently available in commercial products. Building such systems is a very active area of computer science research; although network database systems with many of the features herein described are already on the market, a complete distributed database management system is still a few years away. Our discussion will be of a general nature; in particular, we will not discuss different data models, such as hierarchical, graph-based, and relational, currently used by database management systems.

Database Management Systems
A database is an organized collection of data stored on a computer system. A database management system, abbreviated as DBMS, is a software system for managing and accessing a database. These systems are typically used to manage data that is accessed by applications programs (e.g. packages for statistical analysis), or to answer direct user queries.

It is certainly possible to store data in ordinary files in a computer's file system; in fact, for small datasets that are needed by only a single user, this is often the most practical solution. But for large datasets, and particularly those datasets that are accessed by multiple users, this practice has a number of drawbacks. DBMS', on the other hand, are designed specifically to handle these sorts of situations, and to overcome the limitations presented by using ordinary files.

In particular, DBMS' are designed to handle large datasets that are accessed (often simultaneously) by many users, for both reading and writing. These datasets are usually of critical importance to the user, so protecting the accuracy and integrity of the data stored in them is of paramount importance. Some of the crucial issues that must be dealt with by a DBMS include the following:

* What should be done if more than one person wants to access the data at the same time?

* What happens if one person is changing data at the same time someone else is reading it?

* Is the data safe if the system crashes? What if someone was making changes to the data when it happened?

* How can the data be accessed quickly, even when the dataset is very large?

These are some of the problems that a database manage-
Network Computing
In recent years there has been a strong trend toward replacing mainframes with networks of smaller computers. The most common such arrangement is a local- or wide-area network that connects a collection of personal computers or workstations. This migration is largely the result of the more favorable price/performance ratios currently offered by PCs and workstations. The lower cost is partly due to the high servicing expenses (usually provided by the vendor) required by most mainframes; PC maintenance can frequently be handled by the customer herself. The main reason, however, is the low prices resulting from the intense competition among PC and workstation manufacturers. Furthermore, the explosion in the number of these machines now in use has led to the availability of a wide range of application software to run on them. The proliferation of these smaller computers has put computing power and data closer to their end-users, often right on their desktops.

Another factor contributing to the spread of network computing environments is the growing popularity of the UNIX operating system. UNIX is particularly well-suited for networking, so its increased use has encouraged the implementation of network computing environments.

Unlike the case with mainframes, data storage and processing are distributed in a network environment. Data can be stored at many different sites on the network, and computation can be performed on the machines best-suited to the individual tasks. Network file systems can make data location-transparent to users, so that they don’t have to know where the data they’re using actually resides. Through the use of remote procedure calls, the same can be true of computation. Software can be implemented so that users need not know which machine is running their applications.

Distributed Database Systems
The popularity of network computing environments has led to the development of distributed database systems. A distributed database is a collection of logically-related databases that are connected by a communications network. A distributed database management system (or distributed DBMS) is a software system for managing and accessing a distributed database. A distributed database system consists of a distributed database and its associated distributed DBMS. The distributed database system is designed so that it appears to the user to be a single, unified database. Typically, the data stored by a distributed database system is spread across a number of computers on the network. The data is distributed with the two following considerations in mind. First, data should be stored close to the machines that are most likely to run applications that use it, thus minimizing network traffic. Second, the amount of data stored on the different computers should be well-balanced, to even out the data processing load and thus eliminate potential bottlenecks.

A distributed DBMS differs from a network file system in that its data is logically structured and is accessed by means of a high-level software interface. This interface is written so as to control concurrent access to the data and ensure data integrity. The data in a network file system is only organized into files, and can only be retrieved as files. Note also the distinction between “distributed databases” and “distributed processing”. The former refers to spreading databases across two or more computers, while the latter refers to spreading processing across multiple computers. Distributed database systems typically perform distributed processing, but the two terms are not equivalent.

The client-server model
An increasingly popular network database configuration is the client-server architecture. This refers to a network computing environment in which one or more machines function as database servers. These computers store the data and manage all database operations. Other machines, known as clients, run application programs. When a client application needs to read or write to the database, it sends an appropriate request to a server. The server processes the request and returns the result to the client, which then resumes running its application.

It is important to point out that client-server database architectures are not necessarily distributed database systems. One reason is that there need not be more than one server on a network; thus the data is not necessarily distributed. Another reason is that the client-server model does not require location transparency; in this model it is acceptable to require that the user know where her data resides on the network. In a distributed database system, the actual location of the data should be transparent to the user.

Client-server architectures are frequently used in network computing environments, and not just for managing databases. In addition to database servers, individual machines often function as file servers or mail servers for the network, handling client requests for those resources.

Issues
The questions — described above in Section Database Management Systems — that “undistributed” DBMS' must deal with are also important problems for distributed DBMS'. In fact, these problems become more complex in a distributed environment. In addition, new issues arise because of the distributed setting. We discuss some of these issues in this section, and how they are handled by distributed DBMS'.
Concurrency control

Recall that one of the problems that any DBMS must deal with is what to do when more than one user wants to access the same data at the same time. This can be especially troublesome when one (or both) of the simultaneous users wishes to change or update the data in the database. In the database world, these questions fall under the heading of [som concurrency control]. As its name suggests, concurrency control means managing concurrent access to data by multiple users. Two techniques commonly used by DBMS's to provide concurrency control are data replication and locking.

In one common data replication scheme, each user wishing to access a particular set of data in the database is given her own copy of the data. (The copy is transparent to the user; to her it appears as though she's working with the database itself.) If she only wishes to read data, then she only has to refer to her own copy. But if she wishes to write to the database, all replicated copies must be updated; otherwise, users who are also working with that part of the database will no longer have an up-to-date copy once she has made her changes. Thus each time a user updates the database, all copies of that part of the data that are simultaneously in use must also be updated. This requires a lot of disk writes, which is a slow operation. Thus system performance can suffer under this sort of scheme. In a distributed DBMS, the updates to all replicated data will cause an increase in network traffic, in addition to disk operations. Thus the performance degradation becomes even more of an issue in a distributed DBMS.

An alternative technique often used to enforce concurrency control is locking. When an application program (or a user making direct queries to the database) needs access to a particular piece of data, it requests a lock — a guarantee of temporary exclusive access — to that part of the database. If another process has already been granted a lock on that data, then the application program must wait until that lock has been released. Thus at any time, at most one process has access to any piece of data in the database.

Locking has drawbacks also. The most obvious one is that when multiple processes need access to the same data, all but one of them must wait until they can obtain a lock. This problem can be alleviated somewhat by shrinking the granularity of the lock; that is, by making the lock apply only to a very specific piece of data. This makes conflicts less likely, but also increases system overhead, since processes will have to request locks more frequently. A similar tradeoff takes place when data replication is used. If the granularity of the replicated data is large, then the user needs to request additional data copies less frequently. However, this enhances the probability that she's working with data that is out-of-date. Reducing the granularity, on the other hand, increases the number of data retrievals required, and thus the system overhead.

Another problem introduced by locking is the possibility of deadlock. Suppose that two processes, P_1 and P_2, are running simultaneously, and both need access to both of the data items d_1 and d_2. Suppose further that P_1 requests and is granted a lock on d_1, while at the same time P_2 requests and is granted a lock on d_2. Now neither process can proceed, since each is waiting for the other to release its lock. This situation is called deadlock, and is clearly something to be avoided. DBMS typically have subroutines that periodically check for this sort of condition; if it's found, one of the deadlocked processes is forced to relinquish its lock, so that the other process can proceed.

Deadlock detection is more difficult in a distributed DBMS, since the locks that the processes are competing for may be at two different sites on the network. Since each site typically handles the locks on its own data, it's possible that neither of the two sites realizes that it's waiting for a lock to be released on the other machine. This stalemate situation is known as global deadlock, and is much harder to detect, since there is no central program managing all of the locks. Distributed DBMS' often detect it by "timeout": Once a certain period of time has elapsed without the locks being granted, a global deadlock condition is assumed to exist.

Concurrency control in distributed DBMS' is further complicated by the possibility of communication or site failure. A message relinquishing a lock may be garbled or lost, or a computer may crash without releasing its locks. Either of these situations results in dangling locks, which are no longer needed but have not been released. Distributed DBMS' must have contingency plans for detecting and dealing with these situations.

As an example of concurrency control, consider an airline reservation system. This type of system is centered on a large database that can be accessed by thousands of ticket agents around the world. Clearly some sort of concurrency control is needed to prevent two agents from simultaneously booking the same seat. When one agent tries to reserve a seat, she is granted a lock on that seat. The lock granularity should allow the agent to simultaneously lock two adjacent seats for a couple traveling together, but also allow other agents to book other seats at the same time.

Consistency

Maintaining database consistency is another important issue. Consider a bank, and two customers (A and B) who have accounts there. Suppose that A writes a check for $100 to B, who deposits it in her account. Recording this
transaction in the database thus requires two operations: The balance in A 's account must be decreased by $100 and the balance in B's account must be increased by the same amount. But what happens if an accounting program reads the balances after A's account has been adjusted but not B's? The accounting program will be told that there is $100 less in the bank than is actually the case. The database will be in an inconsistent state.

To prevent this sort of inconsistency, DBMS' allow users to group database operations into transactions. These are sequences of database write operations that are treated as a single unit; either all of the operations in the transaction will be carried out (or committed), or else none of them will be. Furthermore, no other changes will be made to the database between the time the first operation in the transaction is executed and the time that last operation in the transaction is executed. In the banking example, the DBMS would guarantee that the accounting program could not gain access to the account information until after the adjustments to both of the accounts were made. The accounting program would thus see a consistent version of the database.

In a distributed DBMS it is important that, in transactions that affect data at multiple sites, either all of the sites are updated or else none of them are. This is usually ensured through the use of a two-phase commit protocol. In the first phase, the machine initiating the transaction sends a message to all of the sites that will be affected by the transaction, asking them to verify that they are prepared to commit the transaction. If each of these sites responds positively, then the initial machine sends a second message instructing the other sites to actually commit the transaction. If not all of the sites respond positively, perhaps because of a communication failure, then the initial machine sends a second message cancelling the transaction at all of the sites. This ensures that each site maintains a consistent version of the database. This type of protocol requires a significant amount of system overhead.

Disaster recovery
A very important issue is how to protect the integrity of the data during a system crash. Normally, if the system goes down in between transactions, there is no major problem. This is because at the end of a transaction the database will be in a consistent state (assuming that transactions are managed as described above). Of greater concern is the possibility that the system could crash in the middle of a transaction. In this case, some of the changes made inside the transaction may have been executed (i.e., written to disk), but not others. Thus at the time of the crash the database could be in an inconsistent state. In the example above, if the system failure occurred after A's account was debited but before B's account was credited, then the database would understated the bank's deposits by $100.

DBMS' frequently address this problem, known as disaster recovery, by keeping track of all transactions in a log file. Then, in the event of a system crash, the log file is read automatically by the DBMS (once the system is functional again) to see if any transactions were in progress at the time of the failure. If there was such a transaction, all operations in that transaction that were executed prior to the crash are "undone", so that the database is returned to the (consistent) state it was in at the time the transaction started. In a distributed database system each site typically maintains its own log file. Another common technique for guarding against system failure is to make new copies of the disk pages where the data to be updated resides. The transaction operations are performed on the copy. When the transaction is completed, the updated disk page can be remapped to replace the old data in the database in a single, atomic operation.

Query optimization
A primary goal of all DBMS' is to provide fast access to information in the database, even when the database is very large. The speed with which a DBMS can answer a query from a user or an application program relies to a considerable extent on the order in which it carries out the database operations necessary to extract the requested information. Thus to ensure efficient performance, a DBMS must be able to optimize the execution of queries.

As a (rather simplistic) example of the importance of ordering the operations wisely, consider the "Irish presidents" problem. Suppose there is a database containing information on many thousands of current and past US citizens, and a request is made for a list of all people in the database who were both of Irish ancestry and American presidents. One way to satisfy the request would be to first retrieve all people in the database of Irish ancestry, and then select from this (very large) list those people who were also presidents. A much more efficient way to generate the list would be to initially retrieve all US presidents, and then select from this much shorter list those who are also of Irish descent.

Many DBMS' allow precompilation of queries. That is, if a certain type of query will be executed many times, the user can compile it into an optimized form that can then be used for all subsequent invocations; optimization is thus performed only once. For ad hoc queries that will only be executed once, the DBMS must perform the optimization when the query is actually made. Since optimization can be a time-consuming operation, a tradeoff can arise between the time needed to perform the optimization and the time saved by executing an optimized version of the query. In these situations it may be best to perform less extensive (and thus less time-
Query optimization is especially important for distributed databases, particularly since individual queries may involve data stored at more than one site. Communications overhead is a major concern in distributed environments, due to the relatively slow speed of network communication relative to most other operations. Thus it is important to optimize queries so as to minimize the amount of network traffic, in terms of both the frequency of communication and the size of the messages passed. Global optimization, which takes into account communication times between sites, is needed for maximum efficiency, rather than just local optimization at each of the database sites. Note that the more autonomous the individual sites are, the harder this will be to do, since effective global optimization requires that much information on the distributed data be available in a central location.

**Data distribution** A key aspect of the definition of a distributed database system is that the data is distributed among multiple sites on the network. The way in which the data is distributed can dramatically affect system performance.

The question of how data is to be distributed can be divided into two parts, **fragmentation** and **allocation**. Fragmentation refers to how the data is broken into pieces. Once this has been determined, the data fragments must be allocated to various sites on the network. When large networks and databases are involved, finding an optimal (or near-optimal) fragmentation and allocation can be a very difficult problem. While it's important to put data close to users, it's also crucial to distribute the data so as to reduce network traffic, and to balance the distribution in order minimize the chances of bottlenecks appearing. Thus the frequency of access to the data fragments must be considered. Also, the structure of anticipated queries should be taken into account, with the goal of reducing the number of queries requiring data from multiple sites. A sound distribution design strategy should take into account all of these issues.

In order to improve performance, the system designer may wish to store multiple copies of some data fragments, particularly the most heavily-used data. This can result in frequently-accessed data being stored near many or all of its users, as well as reducing contention for individual copies of this data. The cost of such a strategy is that this may incur substantial system overhead; updates must be made to all copies of the replicated data, resulting in more network traffic and additional concurrency concerns. The replicated data fragments will also require more disk space.

Another important feature of a distributed database system is that the data should be **location-transparent**. That is, no matter how the data is fragmented and allocated, the user shouldn't have to know how or where it is stored. The interface to the distributed DBMS should be such that a user, or an application program that interacts with the database, views the distributed DBMS as a single logical database. Note that location-transparency implies that when the data in the database is moved around or restructured, existing application programs won't have to be altered to adjust for the changes.

**Heterogeneous networks** A practical consideration of some importance is that distributed DBMS's should be able to work on heterogeneous networks. Here "heterogeneous" refers not only to computer hardware, but also to the different types of network hardware, operating systems, communication protocols, and even database management systems that are commonly encountered. The last of these is critical, since much of the cost-effectiveness and usefulness of a distributed DBMS may result from its ability to link together a number of existing DBMS's.

A distributed DBMS that is able to run on a wide variety of systems enables widespread sharing of data among databases in environments like universities, where different types of computer systems abound. Another advantage to heterogeneity is that if a distributed DBMS runs on many types of systems, then it's easier to add existing databases to it. This way system administrators can protect their investments in existing systems by being able to integrate them into a larger system, rather than having to replace them.

Hardware and protocol heterogeneity can be achieved through the use of low-level communication interfaces called **gateways**. Once these interfaces have been established, there can still be communications problems if the distributed database system links together different types of DBMS's. Thus it may be necessary to translate between the two (or more) different DBMS's query languages. The software programs that translate DBMS requests into alternative query languages and send them to the appropriate sites are, confusingly, also known as gateways. Note that in a distributed database system with many different types of machines, protocols, and DBMS's, the number of gateways (of both types) required can be very large. This problem could be ameliorated by the adoption of industry-wide standards for such things as data models, query languages, and concurrency protocols. Such comprehensive standards, however, seem unlikely to be established in the near future.

**Advantages and Disadvantages**

As might be expected, there are both advantages and disadvantages to distributed database systems. Perhaps the most obvious advantage is that such systems facilitate...
the sharing of data among large communities of users — for example, among the faculties of different departments in a university — using existing, possibly heterogeneous, computer networks. Thus more users can have access to more data, without having to know where or how the data is actually stored.

**User interfaces** One advantage of a distributed DBMS that can be very apparent to end users is the superior variety of user interfaces available on PCs and workstations as compared with most mainframes. Graphical user interfaces (GUIs), allowing multiple windows and (often) bit-mapped displays, are commonly available on these smaller machines, and greatly enhance the enjoyment and productivity of the user. In a distributed database system, the user can work with a GUI to access data stored elsewhere on the network without having to learn and use the less user-friendly style of command interface that still exists on most mainframes.

**Performance** A related advantage is that computation-intensive applications can be moved off of the database server machine(s) and onto the users’ individual PCs and workstations. This takes some of the processing load off of the servers, thus allowing all users faster access to data. In a mainframe environment, user applications compete with the DBMS software for the computer’s CPU. But by processing the data locally in a distributed environment, greater processing capacity is achieved by keeping many machines busy.

Another way that performance gains can be realized in a carefully designed distributed database system is by moving data closer to the people who use it. By distributing data on the PCs or workstations of those most likely to use it, not only do those users benefit from faster data access, but other users benefit as well, due to the resultant lightening of the load on the other database servers on the network.

Note that, in addition to offering additional functionality such as concurrency control and data consistency, a distributed DBMS may also achieve better performance than network file systems in certain applications. This is because distributed DBMS’ respond to queries, and thus need only send over the network the data that satisfies the specific query. In a network file system, however, only files can be transferred across the network. Thus a much larger amount of data than is actually needed by the requester is likely to be sent, resulting in increased communication time.

**Incremental growth** Distributed database systems also facilitate the incremental growth of databases. New machines, perhaps with new datasets mounted on their file systems, can be incorporated one by one into a distributed DBMS. Thus as the data to be stored outgrows the existing systems, new machines can be added to expand capacity. In a mainframe environment this type of incremental growth is generally not possible; the entire DBMS would have to be replaced with a new system, a much more expensive solution. A related advantage of distributed DBMS’ is that they are well-suited for handling ‘em legacy systems. A legacy is a software program that, although today it might not be the best choice for its job, is so firmly entrenched in the user community (because of years of use, hundreds of applications that invoke it, etc.) that it would not be feasible to replace it. A legacy database system can be incorporated into a distributed DBMS by making it one node in the system. Applications requiring data from that DBMS could still use it, while other programs and users could use data stored on other machines on the network.

**Reliability** Robustness and reliability are other areas in which distributed DBMS’ display advantages. In most cases, a failure at one or more sites on the network will not crash the entire distributed DBMS. In a well-designed distributed DBMS, not only the data but also the control over query processing, concurrency, and disaster recovery is distributed. Thus failure at one point will not render the entire database system useless. Although some data may be temporarily unreachable in the event of such a failure, much of the data should still be accessible. In addition, if the system is designed with careful replication of data on multiple sites, all users may be able to continue working with no ill effects if part of the system goes down.

**Disadvantages** The most telling drawback of distributed DBMS’ is probably the increased complexity of administering and maintaining the database. Instead of just managing a single database, the database manager must now also contend with the network, communications software, data that is replicated on multiple machines, and the backup and recovery of distributed data. There are more possible points of failure, including the machine requesting data, the network, and the machine hosting the data. Testing new applications is harder, since there may be many different combinations of client and server machines that users will want to run the applications on. Software updates are more of a chore; instead of installing an update on a single machine, database administrators must ensure that all machines in the distributed database system are running up-to-date software. Finally, maintaining data security will be more difficult. There is an inherent conflict between granting wider data access to enable many machines on the network to use the database, and restricting access to sensitive data. In environments where sensitive data is stored, a careful balance must be struck between these competing interests.

Another potential disadvantage is that poorly imple-
mented distributed DBMS’ may exhibit worse performance than their centralized counterparts. A system with a very high rate of transactions, and data that is not distributed efficiently, could result in very heavy network traffic. This traffic, combined with the overhead of the software managing the distributed DBMS’ network communications, could cause communication delays that overcome the expected performance gains described above, and result in unsatisfactory performance. Thus careful thought must be given as to how data is distributed and replicated among the machines on the network.

Concluding Remarks
Although fully-functional distributed database systems, as described in this paper, are not yet a commercial reality, they appear to be a promising means of handling large shared datasets in the near future. Systems with many of the capabilities discussed are now available, and distributed client-server databases have been installed at a number of sites. Distributed DBMS’ take advantage of many of the features that have made network computing environments increasingly popular. They distribute the processing load, move the data closer to the people who work with it, and allow cheap, incremental system evolution.

One unknown aspect of distributed DBMS’ is how well the algorithms and protocols they use, such as two-phase commit, will scale up as networks grow larger and connect more and more computers. Further research is also needed on data distribution strategies and on improving transaction management and query processing in a distributed environment. In spite of these obstacles, however, the future of distributed database systems seems bright, aided by the continuing growth in popularity of network computing environments.

References


2. In the computer science literature, “data” is invariably treated as singular, rather than plural. This convention will be followed in this paper.)

3. Workstation” here refers to a desktop computer, usually intended for single user operation, that features a faster processor, more memory, and a larger screen than a PC. Most workstations are UNIX-based and offer elaborate graphical user interfaces.

4. This is known as the “read one, write all”, or “ROWA”, protocol.
Catching Reader Responses on the Fly

by Rosanne G. Potter
Department of English
Iowa State University
Ames, Iowa 50011

Literary criticism, the art of making discriminating judgments or evaluations of literary works, depends in large measure on perspective. Reader Response Criticism approaches literary works from the perspective not of the writer, or the period, or the style, but from the perspective of the reader. One looks at interpretations of literary works and questions the knowledge base, interests, and psychological defenses of the readers who devised the interpretation. For more than 10 years, I have been using computational and statistical means to study linguistic and stylistic features in texts in an attempt to find and quantify textual controls in dramatic literature.

I conducted my first pragmatic study of real readers in the early 1980s but, at that time, did not concern myself with gender. In 1990 when I agreed to chair the Women’s Studies Program at Iowa State, I began noticing differences between male and female student readers in my classes; I wondered whether males and females reacted to different cues or responded to the same cues differently and, most important, how I could catch the responses as the readers were reading. This paper describes my investigation of gendered reading; it begins with a brief summary of other pragmatic studies of readers - noting the presence, or absence, of sex and gender as factors in these studies.

Reader Response Criticism has been the least developed kind of criticism because, until quite recently, readers were thought to make only one contribution to the critical process: they were the sources of error, private associations, and misreadings. These negative judgments of readers flow from the assumption that there really is a “right” or “most complete” reading, and that, if properly trained, all readers can achieve it. This “right reading” assumption has undermined practical critics from I. A. Richards to Elizabeth Flynn and is still very much in practice in most American classrooms.

Although theorists like Wolfgang Iser have written book after book about “the reader,” the researchers who have attempted pragmatic studies of real readers have been few and their methods painfully unsystematic. The first, and most famous, empirical study was performed at Cambridge University by I. A. Richards in the twenties. Richards’ results were so devastating to him, and to many other, that for fifty years after, no one attempted to study the reading skills being learned in literature classrooms. In his 1929 book Practical Criticism, Richards reported the results of asking Cambridge undergraduates to read thirteen poems (authorship not identified and ranging from John Donne to minor poets), then to “comment freely” on their “readings.” Richards documents — in excruciating detail — the many ways of misconstruing meaning in poetry when it is presented “without any hint of provenance” (5). Richards presents his findings not to indict the “products of the most expensive kind of education” (292), but to demonstrate that in all types of educational settings “we must cease to regard misinterpretation as an unlucky accident. We must treat it as the normal and probable event.” (315) He rightly ascribed their poor readings to “bewilderment” (296) caused by the lack of “clues [about] authorship, period, school, the sanction of an anthology, or the hint of a context.” (296)

In her 1978 book The Reader, the Text, and the Poem: The Transactional Theory of the Literary Work, Louise Rosenblatt reported on over twenty years of collecting student responses to unidentified poems; as I see it, she, like Richards, created abnormal test situations. In everyday reading, readers know the name of the author and can easily discover her/his dates, nationality, and school, and may have had their responses shaped by earlier readings of other works by the same writer or by earlier teachings about the writer or the work. By forcing readers to respond to the words of the text only, both researchers deny a reality condition in trying to create an unbiased test situation. This experimental design inevitably sets up perfect conditions for “errors” and encourages the discovery of differences between readers’ responses.

Fifty years after the publication of Practical Criticism, Norman N. Holland in 5 Readers Reading decided to conduct “more or less undirected interviews with a few readers who had taken standard personality tests” (x). He taped extensive interviews with five undergraduate students on Faulkner’s “A Rose for Emily” and then read their responses to the story in light of their “identity themes” (56-62). Instead of finding a great deal of overlap between the readers, Holland found that the readers perceived very different stories. Holland arrived
at "four principles that describe the inner dynamics of the reading experience: "Style Seeks Itself" (113), "Defenses Must Be Matched" (115), "Fantasy Projects Fantasy" (117), and "Character Transforms Character" (121). These principles are psychological descriptions of how readers transform the characters and events in stories to defend their own identity themes while reading fictions.

Holland, like Richards, used a free-response method, but Holland's method was molded by three interventions Richards had not allowed. The readers knew the author and the name of the story (some had even read it before in other classes). The readers' responses and personalities were both elicited by questions from Holland and reported by him. This method of gathering reader responses produces, in Holland's words, no "uniform core" of meaning "from the text" as opposed to "individual variations" contributed "from the people" (366). Thus, while Richards is distressed at the general decrease in ability "to make plain sense of poetry" (12), Holland explains "the way readers respond to literary characters as if they were real people" (xii) by applying Freudian terms ("transformations," "defenses," "fantasies") to the reading experience.

Although they perceive the outcomes of their experiments very differently, neither Richards nor Holland has a model that can be replicated, because neither has an experimental design with clear-cut categories for grouping responses. They inevitably emphasize differences among readers because each reader is treated as a separate case rather than has having features that can be clustered with other similar reader responses. In any study where readers respond "freely" and no attempt is made to identify features in the responses which correlate with features in the texts read, the results will inevitably emphasize difference.

My 1982 study of reader responses to the characters created in the first acts of 21 modern English-language plays<footnote text>could focus on agreement among readers because it asked all readers to respond to the same questions about seven character traits (dominance, intellect, extatability, speculation, poetry, education, attitude), using the same scale (e.g., markedly dominant, moderately dominant, neither dominant nor dominated, moderately dominated, markedly dominated). The reader responses were correlated to features in the language assigned those characters (like high or low use of questions, imperatives, fragments, exclamations, and seven other syntactic and/or semantic features). Since the research design asked specific questions and correlated the results with countable features in texts, there was no difficulty either in finding areas of statistically robust agreement among readers on character traits or in regressing the character trait data against the linguistic features to show which features figured at what levels in readers' judgments of character traits.

Literary scholars who do not know about dependent and independent variables or about objective methods of handling data, and who never attempt to gather qualitative information in quantifiable ways, are destined to discover, as Holland did, responses that have "nothing in common" (366).

Up through Holland, no particular interest in sex or gender as variables shows up. If women participate in the studies, that fact is either not noted or not considered a significant enough factor to require any balancing of the groups being tested. In the 1980s, the idea of gender-balanced samples begins to emerge, but since the general research methods continue to be highly subjective, the introduction of this variable hardly matters; the presence of sex as possible variant does not change the basic methods of analysis.

In Elizabeth Flynn and Patrocinio Schweickart's important 1986 book Gender and Reading, David Bleich reports on conducting an admittedly unscientific study of four females and four males (one of whom was himself) in an attempt to discover differences in male and female ways of reading male and female writers. Bleich's general conclusions are that males and females respond in similar ways to lyric poetry by male and female writers, but very differently to fictions. Men conduct a dialogue with the author about the characters and situations, while women enter the fictional world and allow themselves to see feelings more quickly than men do. This graduate-student pilot study established the theses to be tested in a larger, apparently scientific, contrast of the retellings of Faulkner's short story "Barn Burning" by 00 males and 00 females. Bleich, whose most famous book is entitled Subjective Criticism, does not use objective methods for handling the responses he collected. Instead he reads the responses subjectively and finds, not surprisingly, that they not only confirm his earlier findings, but also allow him to go on to even larger generalizations. The narrator's voice is the "mother tongue" and since separation from the mother is less significant for women than for men, men perceive men as "less-other" than men perceive women.

These assertions may be true, but we should not be misled into thinking that the use of objective methods to collect data means that Bleich's generalizations have any more truth value than if he had arrived at them without consulting any readers. Actually the data he collected and the conclusions are structurally unrelated. Starting from a gender-balanced sample does not necessarily prove anything about gender. Skill in research design is not widely distributed, especially among literary critics who have no training in even recognizing a well designed project when they read it.
In her essay on “Gender and Reading” in the book of the same title, Elizabeth Flynn describes collecting large numbers of responses from males and females (even though that was quite inconvenient at a mostly male school) and using random selection to achieve a balanced number of responses. Unfortunately, she also applies no objective tests to the data so carefully assembled. Like the male critics who preceded her, Flynn reads the essays and judges their adequacy as readings against her own critical standards supplemented by psychological terms describing human interactions. Examples of “domination” by or “submission” to texts are quoted and dismissed in favor of a “third possibility” in which “the reader learns from the experience without losing critical distance; reader and text interact with a degree of mutuality” (270). Flynn concludes by asserting that many males react to disturbing stories by “rejecting” them in an attempt to “dominate” the text, and females “more often arrive at meaningful interpretations of stories because they more frequently break free of submissive entanglement in texts” (285). Like Holland’s, Flynn’s conclusions are interesting and both sets of insights may have, as Oscar Wilde quipped, “the minor merit of being true” but neither researcher has begun to prove anything. They have simply used a new method of establishing “authority” or ethical appeal.

At this point, I wish I could say and here, tah dah, is the reader response study that does what none of the others have done, but I come more to discuss the theoretical issues that are at stake when gender appears in empirical studies of reader’s responses to literature, than to make final report on research.

Since my work on male and female responses to modern British literature texts is still under development and has only been described in print in a Belgian journal, I will describe it here in some detail. The study grew out of two occurrences in a 1990 Modern British literature course in which the students wrote reader responses to each assignment before class discussion. The first reader response, to Chapters 1 through 3 of Oscar Wilde’s The Picture of Dorian Gray, surprised me because the readers’ responses seemed to be sex-linked. Both male and female students commented on the “flowery” language and the ornate tone of the writing, but the males then asked, and I quote, “Is this guy queer?” or asserted “This guy must be a homosexual,” while the females said “He is very sensitive” or “very poetic.” In class, we discussed what they were reacting to and why some males drew conclusions that no females did. I gave the students the “sexual facts” on Wilde: that he was a married man and, as far as his biographers can discern, had not had any homosexual experiences at the time of the writing of this novel; that he did subsequently have such experiences and, five years later, was imprisoned for “gross acts of indecency,” the 1890s’ code term for homosexuality.

A month or so later, the second incident occurred. The students had read their first selection from Bernard Shaw’s An Intelligent Woman’s Guide to Socialism, Capitalism, etc. (Chapters 7-12). Here the difference was much more pronounced and intense: the female students related very positively to the text; they felt that Shaw understood the realities of women’s lives and was arguing for improved economic conditions for women. A strong majority of the male students felt that Shaw was condescending to women and treating them as little better than children. This difference erupted into a full-scale classroom confrontation and to the discovery that the males and females were in some cases using the same passages to prove their different positions. The strongest male and female speakers both wrote papers in support of their readings.

The female, who was simultaneously attending a senior seminar “Language and Gender,” designed a questionnaire using selected passages (ones that “proved” her point, ones that “proved” the male student’s, and ones that both asserted “proved” their opposing positions) and administered it to male and female student friends. Although the sample was small and neither stratified nor random, the results provided more anecdotal support for the proposition that male and female readers drew opposing conclusions from their readings of the selected passages. As a result of these two cases of striking gender differences in reader responses to literary texts, I designed an interactive reading experiment for use during the next offering of the same course.

The project I designed in 1990 and ran in 1991 goes back conceptually to Michael Riffaterre’s 1959 insight (“Criteria for Style Analysis” Word) about the existence of places in literary texts that are commented on by almost all average readers (or ARs). Riffaterre asserts that disagreements (among critics about what passages mean) proves that stylistic devices (or SDs) have surprised readers. The unexpected use of language, according to Riffaterre, elicits interpretation. I wanted to catch ARs responses to SDs in their first reading by inducing them to read new texts on a computer screen and to respond to anything that they found “surprising or unexpected.” My assumption about the best way to get a response (without interrupting the reading process) was to ask readers to take a simple action while reading, e.g., to “double-click” on words that seemed surprising or unexpected.

There were six readings: balanced by gender of writer (three females, three males) and balanced by genre within each gender (two works of fiction, one of non-fiction). The males (Oscar Wilde, Bernard Shaw, and James Joyce) were all from Dublin, though from differ-
ent classes and social spheres; the females (Katherine Mansfield, Enid Bagnold, and Virginia Woolf), were from London and Wellington, New Zealand, were all upper-middle class.

The students self-selected into either MAC-lab readers and control-group readers. The MAC-lab readers came to the lab, entered a few demographic facts (their sex, age, major, and home state) in a logon procedure, and read for the first time an initial segment (chapter, or part of chapter) or a complete short work. As they read, they double-clicked on words, thus highlighting them and (although this was not spelled out to the students), simultaneously moving each highlighted word into a list tagged with the student’s demographic facts.

After completing the reading, the students were asked to reply to six forced-choice post-reading questions designed to establish whether they (1) enjoyed the reading (i.e., did they want to read more by this author), (2) were experienced readers of texts like this one, and (3) felt competent (in terms of vocabulary and general comprehension) in the face of this text. They were then asked four expansion questions (“What is this work about?” “What do you think of the writer?” “What do you notice as repeated?” etc.) and finally, they were asked to write a short paragraph of reaction to the reading. The control group read the same assignment before class, discussed it in small groups, and then wrote about it again afterwards. A contrast of the discursive writings by the MAC-lab and the control group students was anticipated but, unfortunately, not performed in the pilot stage.

The Mac-lab readers were overwhelmingly female (18 to 6). The ratio of females to males in this course is routinely 3 to 1; the same ratio self-selected into the experimental group. The female/male imbalance was intensified when four males completed only five of the six readings, and one completed only four. I had hoped to find that males and females responded to many of the same words, and that some words were more surprising to females than males and vice versa. The results were so skewed as to be statistically unreportable; the best that could be said for the six lists of “surprising words selected” was that they showed as much variance within sex as between sexes.

This study of reader’ responses, especially when it is seen in the context of other pragmatic and theoretical approaches to reading comprehension, can be refined in a number of ways. First, it needs to be conducted on larger samples of males and females which, as the research reported on in this paper shows, means moving out of the small upper-division classroom and into the large, always available, Freshman English pool. Second, more needs to be known about the readers; knowing their sex is not sufficient. If readers could be arranged along a scale of more or less “masculine” or “feminine,” their responses could grouped to see if the social construct of gender is more useful than the biological differentiation into male and female. Third, the whole question of post-reading questions needs thorough re-thinking. Many students reported having very little memory of the texts they read on the MAC screens. The students probably experienced some test and time anxiety because they knew they would have to answer questions after the reading; these anxieties may have interfered with their responses, their comprehension and subsequent memory of the texts.

Possibly, the most important re-design would be to pre-segment the texts, so that passages, rather than words, would identified by clicks. Word orientation tended to mean that “difficult” vocabulary items (including British spellings or usages) dominated the word lists. Passage orientation would group responses so that students who respond more slowly (at the end of a striking passage rather than near the beginning) would still be counted as responding to the same stimulus as the quicker, more experienced readers.

This first of these improvements flows directly out of the comparison between the small and larger studies; it is prima facia clear that if one wants to study male versus female readers, the numbers of males and females need to be larger and more balanced. (Bleich’s report on four subjects is used merely to investigate gender differences; the second study of 00 students is the one that is supposed to convince.) According to Mack Shelley, the statistician I worked with on the pilot stage of this project, I would need a sample size of at least 260 responses (balanced between males and females) to have enough degrees of freedom to start getting statistically significant results.

The second improvement has to do with triangulation, a research design achieved in my “Reader Responses and Character Syntax” project: the relation of two countable features through a third. In my 1982 essay, I counted occurrences of syntactic and semantic features and correlated them with characters who used these features, through readers’ judgments of those characters’ personality traits. Here, I counted the words and the sex of the reader, and probably should correlate these through the reader’s scores on a standard test, like the Bem Sex Role Inventory. The Bem scores could be used to arrange the readers along a female/androgynous/male spectrum and might contribute to a better account of within-gender variance.

The third improvement, eliminating the post-reading questions, would keep the readers’ attention focussed on the reading. The students were certainly distracted from selecting words as surprising, because they were concerned about whether they would have enough time to
complete the assignment in the class period. My study of
gendered reading has just begun.

The Future for Studies of Gender in Reader Responses
When Reader Response Criticism is fully articulated, it
will try to understand how readers’ experiences, experi-
cences tied to their class, age, ethnicity, and gender, affect
their interpretations of literary works; it will try to assess
the impact of information (or the lack of it), inclinations
and disinclinations in readers. Unless testing techniques
are defined that categorize kinds of responses (based on
knowledge, personality traits, gender roles), unless reader
demographics: sex, age, ethnicity, education are factored
in, and until all factors are correlated to features in the
texts, reports about the impact of gender on the reading
of literature will be based on assertions, not upon re-
search.

1 Presented at the IASSIST 92 Conference held in

2 In her 1987 The Return of the Reader (London & New
York: Methuen), Elizabeth Freund summarizes the
“vices” as: “carelessness, self-indulgence and senti-
mentality [with] arrogance and obtuseness...not far behind.”

3 Carbondale and Edwardsville: Southern Illinois Press.

4 Holland had enunciated this model in his 1968 work
The Dynamics of Literary Response (New York: Oxford

5 “Reader Responses and Character Syntax” in Comput-
ing and the Humanities, ed. Richard W. Bailey, (Amster-

6 The Writings of Oscar Wilde ed. Isobel Murray, “The
Critic as Artist, Part II,” 278.

7 These responses formed 25% of the student’s grade,
were collected daily and returned at the beginning of
the next class with brief positive reinforcement comments,
i.e., no comments on writing problems, or “wrong” inter-
pretations, just positive notes on insights, thought
processes, and/or expression.

8 In her best selling 1990 book You Just Don’t Under-
stand, Deborah Tannen’s descriptions of the differences
in conversation styles may explain for the differing
responses. Tannen asserts that women’s conversational
style emphasizes establishing community, while men’s
conversation style emphasizes competing for authority.
The women students may feel recognized and valued
when Shaw explains how the economic system takes
advantage of the unpaid labor of wives and mothers. The
men students may feel condescended to, put in the one-
down position, when Shaw assumes the role of the
authority explaining economic relationships to readers
who do not understand the subject.

9 When many commentators mention a passage, regard-
less of whether they say the same things about it, Rif-
faterre asserts that they do so because the passage is
surprising and calls for interpretation.

10 Relying on common language interpretations of these
terms, I chose not to create a stipulative definition.

11 The texts were the first chapter of Enid Bagnold’s
Diary without Dates, “Bliss” by Katherine Mansfield,
and the first chapter of Virginia Woolf’s To the Light-
house, Chapters 1 through 3 of Wilde’s Dorian Gray, the
first fifteen pages of James Joyce’s The Portrait of an
Artist as a Young Man, and Chapters 7 through 10 An
Intelligent Woman... Each file was approximately
fifteen pages long in Microsoft Word.

12 This was the only required part of the post-reading
responses (all others could be skipped); something had to
be entered here to logout.

13 The texts were re-read when they came up in their
normal places in the syllabus.

14 If I choose to pre-segment the texts, that would mean
a complete re-design of the project. A taxonomy, like
the one described by Teresa Snegrove in her essay on
George Eliot (“A Method for the Analysis of the Struc-
ture of Narrative Texts” in Literary and Linguis-
tic Computing 5 [1990]: 221-225), of narrative-mode tags,
possibly supplemented by persuasive-more tags could be
employed to segment and pre-tag the texts; then the
collected responses to any word within a segment could
be accumulated to see whether narrative and/or persua-
sive modes and gender differences correlate. Differences
in segmentation and/or tagging choices might also turn
out to be gender-matched. If a number of male and
female critics segmented and tagged the same texts,
similarities/differences could be recognized and added
into the variables checked in the responses of student
readers. This augmented approach definitely merits
consideration.
Analyzing Nursing Home Characteristics: Issues In Comparing State And Federal Data Sources.

by Joel B. Cowen
Coordinator Health Services Research
College of Medicine at Rockford
University of Illinois

Introduction
Aging of the population is increasing the importance of long-term care and nursing homes as a component of the health care system in the United States. Policy planners need to fully understand the nature of the industry, as well as emerging trends.

Nursing homes, as we know them currently, are a relatively new entity on the health care landscape. The present long-term care system largely emerged in the early 1970s following the impact of Medicare. Earlier facilities like board and care homes had sprouted up after the passage of social security, along with was activity by some churches and fraternal groups who had built retirement centers for their members.

Since nursing homes are only a couple of decades old and changes in regulations and financing impacting them are being seen with increasing frequency, nursing homes could well change in form or format into the next century, when the elderly portion of the population will move toward one-fifth as baby boomers edge into their senior years.

Planning for an essential service like long-term care requires accurate information. Conjecture cannot take the place of data vital for shaping the nature of future nursing homes. However, despite the importance of nursing homes as part of health care, no comprehensive information system is in place nor are basic definitions of data elements accepted in a widespread manner. The quality of planning suffers when information is limited. For instance, the impact on nursing homes of recent changes in Medicare reimbursement on hospitals needs to be known.

In this paper, current data for nursing homes is analyzed. The two major federal data sets for nursing home facility and resident characteristics, plus similar information collected in one major state, Illinois, are reviewed. Data elements obtained and their definitions are compared and, finally, recommendations made as to how to improve the present situation.

Nursing Home Data Sources
National Nursing Home Survey
The National Nursing Home Survey (NNHS) is a continuing series of national sample surveys of nursing homes, their residents and staffs by the National Center for Health Statistics (NCHS). Three surveys have taken place, in 1973-74, 1977 and 1985. No plans have yet been announced for a fourth survey.

The surveys employ a stratified two-stage probability design, first the selection of facilities, then the selection of residents and employees. Data has been collected using both personal interviews and forms for self-completion. In 1985, added information was collected from relatives of residents, the “next of kin” questionnaire. In 1977 and 1985, a sample was also taken of persons discharged from the home in the past year, whether alive or dead.

The NNHS is the only national survey which includes variables for individuals which can be analyzed against each other. Other sources group data, which cannot be crosstabbed or correlated.

All data is available on tape from the National Center for Health Statistics. Figure 1 shows the available variables. Tapes are also available for the states of California, Illinois, Massachusetts, New York and Texas. More cases were provided in these states so that reliable estimates could be made.

Results appear in written form in Vital and Health Statistics, Series 13, Nos. 97, 98, 102, 103, and 115 plus Advance Data, Nos. 131, 135, 142, 147, and 152 from NCHS.
### Summary of 1985 NNHS Data Tapes by Type of File

<table>
<thead>
<tr>
<th>Facility file</th>
<th>Resident file</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility number</td>
<td>Facility number</td>
</tr>
<tr>
<td>Ownership code</td>
<td>Age</td>
</tr>
<tr>
<td>Number of beds (1985 and 1984)</td>
<td>Sex</td>
</tr>
<tr>
<td>Certification status</td>
<td>Race</td>
</tr>
<tr>
<td>Per diem rates by certification status</td>
<td>Hispanic origin</td>
</tr>
<tr>
<td>Admissions (1984)</td>
<td>Marital status at admission and currently</td>
</tr>
<tr>
<td>Residents days (1984)</td>
<td>Living children</td>
</tr>
<tr>
<td>Services offered to residents</td>
<td>Date of last admission</td>
</tr>
<tr>
<td>Services offered to non-residents</td>
<td>Residence before admission</td>
</tr>
<tr>
<td>Physician service arrangements</td>
<td>Hospital stays while a resident</td>
</tr>
<tr>
<td>Full- and part-time staff</td>
<td>Previous nursing home stays</td>
</tr>
<tr>
<td>Part-time staff hours</td>
<td>Diagnoses at admission and currently</td>
</tr>
<tr>
<td>Nursing staff hours</td>
<td>Mental disorders</td>
</tr>
<tr>
<td>Volunteer staff</td>
<td>Therapy services received</td>
</tr>
<tr>
<td>Geographic region recode</td>
<td>Vision and hearing status</td>
</tr>
<tr>
<td>DHHS administrative regions</td>
<td>Activities of daily living</td>
</tr>
<tr>
<td>MSA recode</td>
<td>Adapted instrumental activities daily living</td>
</tr>
<tr>
<td>Facility weight</td>
<td>Behavioral problems</td>
</tr>
</tbody>
</table>

| Record length | 665 | Record length | 873 |
| Block size | 19,950 | Block size | 17,460 |
| Number of records | 1,078 | Number of records | 5,238 |

<table>
<thead>
<tr>
<th>Discharge file</th>
<th>Nursing staff file</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility number</td>
<td>Facility number</td>
</tr>
<tr>
<td>Age at discharge or date of birth</td>
<td>Member of staff or other arrangement</td>
</tr>
<tr>
<td>Sex</td>
<td>Type of position</td>
</tr>
<tr>
<td>Race</td>
<td>Length of work experience</td>
</tr>
<tr>
<td>Hispanic origin</td>
<td>Hours worked</td>
</tr>
<tr>
<td>Marital status at admission and at discharge</td>
<td>Salary</td>
</tr>
<tr>
<td>Date of admission and discharge</td>
<td>Services performed</td>
</tr>
<tr>
<td>Discharge status (alive/dead)</td>
<td>Employment conditions</td>
</tr>
<tr>
<td>Residence before admission</td>
<td>Sex and age</td>
</tr>
<tr>
<td>Residence after discharge for live discharges</td>
<td>Ethnicity</td>
</tr>
<tr>
<td>Hospital stays while a resident</td>
<td>Marital status</td>
</tr>
<tr>
<td>Nursing home stays before and after sample stay</td>
<td>Children living at home</td>
</tr>
<tr>
<td>Diagnoses at admission and at discharge</td>
<td>Education</td>
</tr>
<tr>
<td>Mobility staff</td>
<td>Staff weight</td>
</tr>
<tr>
<td>Continence status</td>
<td></td>
</tr>
<tr>
<td>Sources of payment at admission and at discharge</td>
<td></td>
</tr>
<tr>
<td>Discharge weight</td>
<td></td>
</tr>
</tbody>
</table>

| Record length | 544 | Record length | 307 |
| Block size | 21,760 | Block size | 21,490 |
| Number of records | 6,017 | Number of records | 2,760 |

| Expense file |  |
|---------------|  |
| Facility number |  |
| Expenses and revenues |  |
| Expense weight |  |

| Record length | 366 | Record length | 17 |
| Block size | 18,300 | Block size | 17 |
| Number of records | 731 | Number of records | 17 |
Inventory of Long-Term Care Places

The federal Inventory of Long-Term Care Places (ILTCP), formerly part of the National Master Facility Inventory (NMFI), was conducted in 1986 by the Census Bureau for the National Center for Health Statistics. This differs from the NNHS in that the entire sample of “nursing and related care homes” is covered. Variables are limited to ownership, certification status, number of beds, residents, and race of residents. Information on residents is tabulated in grouped data for reporting.

One way that the ILTCP differs from the NNHS is that special homes for the mentally ill, developmentally disabled, and other groups with special handicaps are included.

Also available on tape from NCHS, the full set of variables is shown in Figure 2. Both the NNHS and ILTCP are also available through the Inter-University Consortium for Political and Social Research.

| Figure 2 |
| SUMMARY OF NATIONAL MASTER FACILITY INVENTORY DATA TAPES BY TYPE OF FACILITY |

<table>
<thead>
<tr>
<th>Hospitals</th>
<th>Nursing homes and other health facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name -</td>
</tr>
<tr>
<td>Name of administrator</td>
<td>Adress -</td>
</tr>
<tr>
<td>Ownership</td>
<td>Number of beds</td>
</tr>
<tr>
<td>Type of facility</td>
<td>Ownership</td>
</tr>
<tr>
<td>Number of beds</td>
<td>Type of facility</td>
</tr>
<tr>
<td>Days of care</td>
<td>Ages served</td>
</tr>
<tr>
<td>Discharge</td>
<td>Sexes served</td>
</tr>
<tr>
<td>Admissions</td>
<td>Number of residents</td>
</tr>
<tr>
<td>Type of service</td>
<td></td>
</tr>
<tr>
<td>Outpatient visits</td>
<td></td>
</tr>
<tr>
<td>Employees</td>
<td></td>
</tr>
<tr>
<td>facilities and service offered</td>
<td></td>
</tr>
<tr>
<td>Record length ..........</td>
<td>840</td>
</tr>
<tr>
<td>Block size ..........</td>
<td>8,400</td>
</tr>
<tr>
<td>Number of records ..........</td>
<td>7,678</td>
</tr>
<tr>
<td>Number of reels ..........</td>
<td>1</td>
</tr>
</tbody>
</table>

Illinois Annual Survey of Long-Term Care Facilities

Most states collect data on their nursing homes and their residents. This is usually made necessary for the state’s licensure certification or planning functions. Illinois is typical. The Illinois Department of Public Health (IDPH) has conducted annual surveys of long-term care facilities since 1981. These surveys are carried out under the state’s Certificate of Need Act which authorizes data collection for planning purposes.

In addition to using the data to create an inventory of long-term care services and bed need plan, IDPH makes survey data available to others who have an interest in long-term care, both public and private. Other state agencies also constitute a major user.

The Illinois survey format is relatively standardized year to year with supplemental “special” studies each year. Nearly a thousand licensed facilities (including mental health/DD) receive the questionnaires and participation is almost universal.

UICOM-R Long-Term Care Report

The Health Services Research office of the University of Illinois College of Medicine at Rockford has created reports characterizing nursing homes in northwest Illinois. These reports, completed in 1980, 1985 and 1990, utilize information from the IDPH Long-Term Care Facility Survey.

This effort is notable because few local area studies are conducted which track changes in the industry on a local area basis. State reports generally do not evaluate long-term changes or focus on regional differences.
Data Elements
In this section, certain data elements have been selected to illustrate the availability of nursing home data, so as to reveal potential sources, their differing methods of gathering, commonalities and differences in definition.

What Is a Nursing Home?
Defining a nursing home is not a clear and simple task. The NNHS includes all types of "nursing homes" regardless of their level of care, participation in Medicare or Medicaid or licensure. No "Board and Care" homes were included or those providing residential care. They define nursing home as:

Facilities with three or more beds that provide to adults who require either nursing care or personal care (such as help with bathing, correspondence, walking, eating, using the toilet, or dressing) and/or supervision over such activities as money management, ambulation, and shopping. Facilities providing care solely to the mentally retarded and mentally ill are excluded. A nursing home may be either freestanding or a distinct unit of a larger facility.

Illinois relies on licensure for its survey definition. In Illinois, a nursing home is defined as a: "private home, institution, building, residence, or any other place which provides personal care, sheltered care or nursing for three or more persons who are not related to the owner."

Long-term care institutions in Illinois are further classified into two types of care, nursing and sheltered care.

Nursing care includes the provision of diagnostic, therapeutic and rehabilitative care under a patient’s plan of care as prescribed by a physician. In addition to the medically oriented care given by nurses and the living assistance given by aides, other services commonly provided by a facility that provides nursing care includes physical therapy, speech therapy, occupational therapy and social activities. There are two levels of nursing care, skilled and intermediate, with the levels differing in the amount of available nursing expertise and supervision provided to residents.

Sheltered care includes the provision of personal care and support in daily activities with limited nursing consultation available, such as the taking of medications.

The term "sheltered care" is somewhat unique to Illinois. Other states tend to utilize the terms "assisted living" or "personal care."

The federal NNHS survey relies on Medicare/Medicaid definitions for designation of skilled or intermediate care levels. Other beds are shown as "not certified."

Both units of government agree in that a facility must have three beds and provide nursing or personal care. Illinois does not consider personal care to determine a nursing home unless there is nursing consultation, such as the taking of medication. Terms common to definitions such as nursing care and personal care are not precisely defined.

Another area for possible confusion is whether facilities such as those for the mentally ill or developmentally disabled are counted. The NNHS excludes these. The ILTCP includes them as does Illinois. The ILTCP, which begins with categories similar to the NNHS, goes beyond these categories to also include homes for unwed mothers, substance abuse, orphans and the terminally ill (hospice).

In general, the definition of nursing home in this country is still imprecise, leaving board and care, congregate living and certain retirement centers and specialized facilities in a zone of uncertainty.

Beds
Beds are classified in various ways such as licensed or unlicensed, set-up and staffed, or occupied. The NNHS and ILTCP both primarily use set-up and staffed, while Illinois uses licensed beds. Licensed beds usually reflect capacity whether actually used or not.

The creation of swing beds and distinct unit skilled nursing beds at hospitals has increased in recent years. Swing beds are those that can be used for either acute or extended care through reclassification of the patient who doesn’t actually move from the bed. Swing beds are certified by the Medicare program. Distinct units provide only extended care. For the most part they are utilized for hospital patients no longer needing acute hospital nursing care but who are not yet
well enough to return home. Under the “DRG” reimbursement system, hospitals benefit financially from discharges to a lower level of service.

The NNHS totally excludes hospital-based beds. On the other hand, the ILTCP includes long-term care units in hospitals. Presumably this includes both swing beds and distinct units. Illinois counts distinct units as a “nursing home” category, but does not consider swing beds in this category.

As is probably obvious, counts of beds can differ greatly according to the counting method. Figure 3 shows some of these conflicting counts.

<table>
<thead>
<tr>
<th>Figure 3</th>
<th>COMPARISON OF BED COUNTS: NNHS, ILTCP, AND IDPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursing Homes in U.S.</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>NNHS</td>
</tr>
<tr>
<td>Nursing Care</td>
<td>14,400</td>
</tr>
<tr>
<td>Hospitals</td>
<td></td>
</tr>
<tr>
<td>Not Certified</td>
<td>4,700</td>
</tr>
<tr>
<td>Residential</td>
<td></td>
</tr>
</tbody>
</table>

| Nursing Homes in Illinois | | |
| Total | ILTCP | IDPH |
| Nursing Care | 744 | 898 |
| Hospital Based | 25 | 734 |
| Residential | 48 | |

**Admission, Discharge and Stay**

Length of stay is important for policy decisions and planning of resources, not to mention actuarial calculations such as those for the long-term care insurance industry. Average length of stay cannot be calculated with ease as is done in hospitals:

\[ \text{Patient Days} = \frac{\text{Average Length}}{\text{Discharges}} \times \text{Stay} \]

Nursing home stays more often cross multiple years and may involve intervening hospital stays.

The ILTCP collects annual admissions and number of “residents last night.” Average length can only be calculated if the assumption is made that “residents last night” times 365 = an estimate of patient days. IDPH collects patient days, admissions and discharges so that an estimate of average length of stay can be calculated. Again, the year to year crossover can be a problem in such calculations.

The NNHS is far more precise in its treatment of stay length. This is important because a good source is needed which differentiates the characteristics of stay types, especially the nature of post-hospital short stays from longer term chronic type stays. The differing nature of different “types” of nursing home residents is important for policy.

The NNHS provides for a question on duration of stay at discharge. Recently they have reconsidered this indicator of stay length in favor of the long-term care use history of individuals. An individual may be admitted and discharged several times a year and, in fact, that pattern has been found to be common.

Additionally, nursing homes treat stay definitions differently, especially with regard to temporary transfers to hospitals. Some consider the movement to be a discharge with formal readmission on return, while others make no such change in
status. Some facilities include the hospital stay as part of the stay, while others exclude them or calculate “bed-hold” days. Illinois provides no instructions on how to deal with the reporting of “bed-hold” days. This is left to the institution.

Another element of admissions and discharges is the place of origin and discharge - Where are residents coming from? and Where are they going? The ILTCP does not collect this, but both the NNHS and Illinois do. The NNHS collects “living arrangement prior to admission and after discharge,” while Illinois asks for “admissions from” and “discharges to.” The categories are similar except that the NNHS is far more precise in terms of home residence. A comparison is shown below in Figure 4.

<table>
<thead>
<tr>
<th>Figure 4</th>
<th>LTC ADMISSIONS AND DISCHARGES COMPARISON OF NNHS AND IDPH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NNHS</td>
</tr>
<tr>
<td>Own Home or Apartment</td>
<td>Relative’s Home or Apartment</td>
</tr>
<tr>
<td>Other Private Home or Apartment</td>
<td>Retirement Home</td>
</tr>
<tr>
<td>Retirement Home</td>
<td>Boarding House, Room</td>
</tr>
<tr>
<td>Other Nursing Home</td>
<td>Other Nursing Home</td>
</tr>
<tr>
<td>General Hospital</td>
<td>Mental Hospital</td>
</tr>
<tr>
<td>Mental Hospital</td>
<td>Chronic Disease</td>
</tr>
<tr>
<td>Hospice</td>
<td>Other</td>
</tr>
</tbody>
</table>

Resident Demographic Characteristics
Recording the demographic characteristics of residents provides an essential component of any description or analysis of the nursing home industry. Age, race, and sex constitute the core of these descriptors.

Birthdate of the resident is collected by the NNHS, allowing any age groupings. The ILTCP utilizes three groups: 0-21, 22-64 and 65+. IDPH provides for seven groups: 0-17, 18-44, 45-59, 60-64, 65-74, 75-84 and 85+. Sex (gender) is collected by the NNHS and IDPH, but not the ILTCP. Nursing homes tend to be dominantly female.

Like many of these data elements, race is classified three different ways. The NNHS applies the usual census format in which race (white, black, American Indian, and Asian) is a separate variable from ethnic origin (Hispanic). Illinois, however, uses combined racial/ethnic groupings (white, non-Hispanic; black, non-Hispanic; Asian, non-Hispanic; and Hispanic). The ILTCP collects only the number of black and Hispanic residents, no other racial groups.

Two resident elements in the NNHS only are marital status (current and at admission) and number of living children.

Health and Activity
Health of the individual is often expressed either through categorization of the major condition or disease resulting in the admission or indications of which activities of daily living (ADLs).

The ILTCP does not record diagnosis at all. Illinois utilizes groupings of ICD-9 codes such as “Diseases of the Circulatory System” or “Musculoskeletal Diseases.” The NNHS lists certain common reasons for nursing home placement such as Stroke, Hip Fracture and Alzheimer’s Disease. The NNHS also collects indicators of health status prior to admission, including DRG if hospitalized.
### Figure 5
DEMOGRAPHIC CHARACTERISTICS OF NURSING HOME RESIDENTS

<table>
<thead>
<tr>
<th>Indicator</th>
<th>NNHS</th>
<th>ILTCP</th>
<th>IDPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Individual</td>
<td>0 - 21</td>
<td>0 - 17</td>
</tr>
<tr>
<td></td>
<td>Years Collected</td>
<td>22 - 64</td>
<td>18 - 44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65+</td>
<td>45 - 59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grouped Only</td>
<td>60 - 64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75 - 84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>85+</td>
</tr>
<tr>
<td>Sex</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Race/Ethnic</td>
<td>White, Black, Amer. Indian, Asian Hispanic</td>
<td>Black, Hispanic Only</td>
<td>White, non-Hisp. Black, non-Hisp. Asian, non-Hisp. Hispanic</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Yes, current and at admission</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Living Children?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

### Figure 6
HEALTH AND ACTIVITY STATUS VARIABLES FOR NURSING HOME RESIDENTS

<table>
<thead>
<tr>
<th>Indicator</th>
<th>NNHS</th>
<th>ILTCP</th>
<th>IDPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activities of daily living (ALDs)</td>
<td>Bathing</td>
<td>Bathing</td>
<td>Bathing</td>
</tr>
<tr>
<td></td>
<td>Dressing</td>
<td>Dressing</td>
<td>Dressing</td>
</tr>
<tr>
<td></td>
<td>Eating</td>
<td>Eating</td>
<td>Eating</td>
</tr>
<tr>
<td></td>
<td>Walking Inside</td>
<td>Walking</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Walking Outside</td>
<td>Medication</td>
<td>Medication</td>
</tr>
<tr>
<td></td>
<td>Toileting</td>
<td>Shopping or</td>
<td>Shopping or</td>
</tr>
<tr>
<td></td>
<td>Transfer</td>
<td>Letter Writing</td>
<td>Letter Writing</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>DRG if Hospitalized</td>
<td>No</td>
<td>15 Groupings of ICD-9</td>
</tr>
<tr>
<td></td>
<td>Disease or Condition Resulting in Admission</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15 Categories such as Hip Fractures; Prior Health Status</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Payment Source
The ways that residents pay for care is an important issue in the delivery of care. The ILTCP does not collect payment source. All sources of payment are indicated on the NNHS, while IDPH obtains only major payment source. The NNHS includes insurance within private pay, while IDPH breaks it out.

Nursing Home Trends
Despite their drawbacks and lack of standardization, the studies reviewed in this report can be used to form a picture of trends in nursing homes and their residents. The focus in this section is on recent changes in the industry.

![Figure 7](image)

**Figure 7**
PAYMENT SOURCE VARIABLES FOR NURSING HOME RESIDENTS

<table>
<thead>
<tr>
<th>NNHS</th>
<th>IDPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicare</td>
<td>Medicare</td>
</tr>
<tr>
<td>Medicaid</td>
<td>Medicaid</td>
</tr>
<tr>
<td>Private (own) Pay</td>
<td>Private (own) Pay</td>
</tr>
<tr>
<td>VA</td>
<td>VA</td>
</tr>
<tr>
<td>State Agency</td>
<td>State Agency</td>
</tr>
<tr>
<td>Other Public Pay</td>
<td>Other Public Pay</td>
</tr>
<tr>
<td>Church, Foundation, Agency</td>
<td>Insurance</td>
</tr>
<tr>
<td>Life Care Funds</td>
<td></td>
</tr>
</tbody>
</table>

ILTCP (1967-1986 unless otherwise noted)

* The number of nursing homes nationally grew 18.2%.

* The number of nursing home beds grew 105.0%.

* The number of nursing home residents grew 106.5%.

* Persons 65+ in nursing homes grew 45.5% from 1967-1976, but then declined 6.2% from 1976 to 1986.

* Occupancy stayed around 92%.

NNHS (1977-85 unless otherwise noted)

* The number of elderly patients discharged from hospitals to nursing homes increased 36% from 1982 to 1985.

* The proportion of nursing homes affiliated with a nursing home chain rose significantly 1977-85 from 28% to 41% of all facilities.

* Discharges rose 9.5%.

* Women in nursing homes aged 85+ rose from 34% to 41% of residents.

* Race and ethnic status was obtained for the first time in 1985. Minorities were underrepresented and generally younger than the white population.

* The proportion of individuals not dependent on help for mobility or continence dropped from 40.1% to 30.1%.

* Medicare-covered days in skilled nursing facilities per 1,000 beneficiaries dropped from 370 in 1977 to 320 in
Residency in a nursing home for persons 85+ dropped from 257 per 1,000 in 1974 to 220 in 1985.

The average length of stay rose from 2.7 to 2.9 years.

The proportion of nursing home residents with mental disorders rose, while circulatory disorders fell.

**UICOM-R (IDPH Data for Northwest Illinois)**

- The ratio of beds per thousand population 65+ fell from 92.9 in 1980 to 85.9 in 1990. The elderly population is increasing more rapidly than the facilities for care. About 6.3% of persons aged 65 years and up reside in area long-term care facilities, down from 7.3% a decade ago.

- The overall annual occupancy of general long-term care facilities in northwest Illinois is 87.3% based on beds licensed. Occupancy rates have increased since 1980 when the corresponding rate was 84.4%.

- Just under half (47.7%) of northwest Illinois long-term care facilities are owned by for-profit concerns, down from 60.0% in 1981. The remainder are owned by churches (18.5%), government (13.9%) and not-for-profit corporations (20.0%).

- Hospitals are increasingly entering the long-term care “business.” Nine of the fifteen hospitals in northwest Illinois have swing beds or distinct units.

- Most general long-term care residents are age 75 or older (81.0%), female (75.2%), and white (97.3%). Residents aged 85 and over made up 47.7% of residents in 1990 but only 39.2% in 1980. The median age is now estimated to be 84.3 years, up from 82.1 in 1981.

- Leading admission diagnoses include the circulatory system (28.6%), nervous system (13.8%), and musculoskeletal (11.8%). One in ten residents is reported to have Alzheimer’s disease. Mental illness, nervous system and musculoskeletal have risen, circulatory has declined.

- Nursing home residents are highly dependent on others for performing certain activities of daily living (ADLs) including bathing (62.8%) and toileting (56.6%). Dependency rates have been increasing.

- Medicaid was the source of payment for 46.4% of the long-term care residents in 1989. Private payers covered 48.9%. Some of the remainder were covered by Medicare (2.7%). Medicaid coverage has been rising.

- The average daily charge for a skilled care double bed in 1990 was $73, up from $48 in 1985. Intermediate care averaged $56 for a double, up from $41 in 1985. Many institutions make additional charges for supplementary services as rehabilitation therapies, bandages, or assistance in bathing.

- 65.2% of area beds are certified for Medicaid residents, but only 6.3% are certified for Medicare. Medicare certification has been stable or declining.

**Improving Nursing Home Data**

**OBRA Data Requirements**

Beginning this year, all nursing homes certified to provide care under Medicare or Medicaid must use assessment instruments required by the state and approved by HCFA. The instrument must include a uniform minimum data set (MDS) of care screening and assessment elements with common definitions. The MDS contains a comprehensive set of data elements which describe most nursing home residents nationally. If modified to add certain elements, the MDS could form the core of a standardized instrument which could be collected and analyzed on a national basis to analyze the characteristics of nursing home residents (see Figure 8).

**Other Sources**

For homes certified by Medicare or Medicaid, financially-centered reports are filed with government or a fiscal intermediary. These “cost reports” generally include a great deal of information on the nature of the facility.
Conclusion
As has been shown in this paper, data describing nursing homes nationally is haphazard at best, and lacking in standardization of definitions. Sources usually cannot be compared to each other, because of differing coverage, variables, and definitions. Of the two federal sources, one is a periodic sample which is relatively comprehensive, which employs several sub-component studies. The other is a periodic census with relatively sparse variables. Neither follows a regular schedule. One was last completed in 1985 and the other in 1986. No immediate plans are currently in place for repeating these studies so as to yield more timely data.

Another data source is the surveys performed by individual states. Most states have annual surveys of the type exemplified by Illinois. States could form the framework for regular assessments of nursing homes as long as guidelines for definition and collection are put forward by the federal government. Much like the vital statistics system operates, states could report to the National Center for Health Statistics, which could compile the information.

Long-term care is too important a component of health care not to have systematic data reporting. Action is needed in this direction.

1 Presented at the IASSIST 92 Conference held in Madison, Wisconsin, U.S.A. May 26 - 29, 1992.
Preliminary announcement & CALL FOR PAPERS

The 1994 IASSIST Conference will feature presentations on a wide variety of topics of interest to social scientists, data archivists, librarians, information specialists, researchers, programmers, government agency administrators, as well as data producers. The Program Committee is now soliciting papers on issues relating to acquisition, preservation and dissemination of machine-readable numeric, text and image data. Papers of special interest would be those focusing on data related issues for research on Global Warming, AIDS, poverty, or other comparative social research themes. Technical topics could include the Uses of the Internet, UNIX Applications in Archives, or Centralized vs Decentralized computing. Library issues could cover Bibliographic Access Tools, Indexing Standards or User Services. Also of special interest would be discussions on solutions to the dilemma raised in an arena of expanding opportunities and diminishing resources.

Conference Location:
Sheraton at Fisherman’s Wharf Hotel
2500 Mason Street
San Francisco, CA. 94133

Conference Site:
The perfect setting for a conference, the Sheraton at Fisherman’s Wharf is located a block from the picturesque sights and sounds of Pier 39, Ghirardelli Square and The Cannery. Downtown shopping, Chinatown and the eclectic North Beach are minutes away via a complimentary hotel limousine. The 4-story building features 525 rooms and spacious suites. Hotel services include concierge assistance, heated pool, car rental/travel arrangements, hair salon, valet and a health club. The hotel is easily reached from San Francisco International Airport.

Hotel and Conference Rates:

<table>
<thead>
<tr>
<th></th>
<th>Member</th>
<th>Non-Member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hotel Rate:</td>
<td>$115</td>
<td>$15</td>
</tr>
<tr>
<td>Conference only</td>
<td>$175</td>
<td>$225</td>
</tr>
<tr>
<td>Conference &amp; Workshop</td>
<td>$225</td>
<td>$275</td>
</tr>
<tr>
<td>Workshop only</td>
<td>$100</td>
<td>$100</td>
</tr>
</tbody>
</table>

Non-member conference rate includes a 1-year membership in IASSIST.
A late fee of $25 will apply to conference or workshop registration received after April 9, 1994.
Local Arrangements Chairs:
Ilona Einowski  
census85@ucbcmsa.berkeley.edu  
Jean Stratford  
Jsstratford@ucdavis.edu

Program Committee Chairs:
Martin Pawlocki  
iyw6pgm@mvs.oac.ucla.edu  
Elizabeth Stephenson  
iyw6pgm@mvs.oac.ucla.edu

CONFERENCE INTENTIONS FORM

Please return this form by December 15, 1993 to:

IASSIST 1994 PROGRAM COMMITTEE  
c/o Pawlocki/Stephenson  
Institute for Social Science Research  
University of California, Los Angeles  
405 Hilgard Avenue  
Los Angeles, CA 90024-1484 U.S.A.

NAME:

TITLE:

AFFILIATION:

MAILING ADDRESS:

 ELECTRONIC MAIL ADDRESS:

TELEPHONE:  FAX:

Below, check all that apply:

☐ I propose to conduct a workshop on the following subject:

☐ I intend to submit a paper on the following topic/title:

☐ I would like to hold a panel/seminar/roundtable discussion on the topic of:

☐ I am interesting in mounting the following post session/display:

☐ Please send conference information to my colleague:

☐ I will be willing to Chair a session.

☐ Please keep me on the mailing list.
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)

---

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)

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

Name / title

Institutional Affiliation

Mailing Address

City

Country / zip / postal code / phone