The Journal of Sociology & Social Welfare

Volume 9
Issue 4 *December*

Article 11

December 1982

Distributed Data Processing: A Timely Approach for Social Welfare Agencies

Sumit Sircar
University of Texas, Arlington

Dick Schoech
University of Texas, Arlington

Lawrence L. Schkade
University of Texas, Arlington

DISTRIBUTED DATA PROCESSING: A TIMELY APPROACH FOR SOCIAL WELFARE AGENCIES

By

Sumit Sircar, Dick Schoech and Lawrence L. Schkade

The University of Texas at Arlington

ABSTRACT

Centralized management oriented information systems in state social welfare systems are evolving. Due to decreasing computer hardware costs, computing power can now be distributed throughout a social welfare system to locations where it meets worker and manager data processing needs most efficiently and effectively. To distribute computing power yet maintain one integrated organizational computing system requires an understanding of distributed data processing (DDP) and its implications for an organization. This article explains the DDP concept through its historical development, illustrates the types of DDP available to an agency, and discusses the major pitfalls in moving into a DDP environment. It also provides guides for implementing and managing DDP systems and discusses the implications of DDP for social welfare workers and managers.

Introduction

Decision making in human services is often complex, and making the best possible decisions requires timely and accurate data and reports. In most human service organizations, data and reports are either processed manually or by a central computing system that is designed to serve the needs of upper management. In the latter case, the computerized system usually requires considerable time and effort by lower organizational units to supply data, while the resulting system reports and summaries are perceived to be of limited value by those units. Typically, the specialized data processing needs of supporting units are not fulfilled.

The nature of the need for computing at the local level is illustrated in the following scenario. Consider a human service worker who is venting some frustration to the unit supervisor. "It seems that half of my time is spent filling out data forms for the central office. After spending all that effort, they call me later to update the data or resubmit it, because they don't trust their computer records. And in spite of all that work which takes time away from clients, the computer still doesn't give me back the information I really need to help my clients."

"What I think this office needs is its own computer. My brother-in-law has a microcomputer in his small business that he uses for customer files, payroll, inventory control and other purposes. This office could use a small computer like his to store and retrieve casework data and to hook up to the big departmental computer, so we can input our data directly into the system instead of sending them manually completed forms. Besides, we could quickly respond to the special data requests the central office frequently makes."

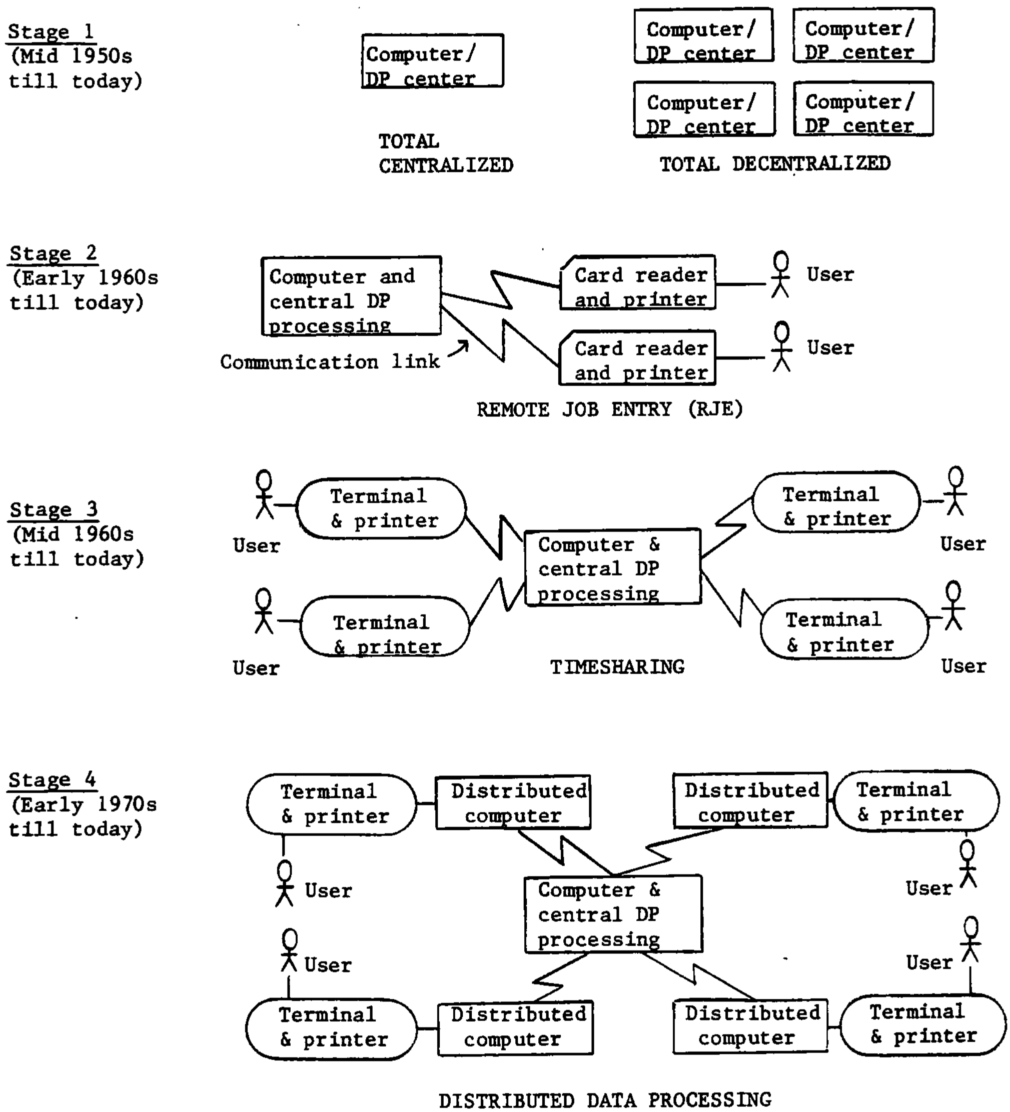
The foregoing scenario, adapted from actual situations observed by the authors, illustrates the need for computing systems that not only serve the specific client data needs of direct service workers in a timely fashion but also supply upper management with data summaries and reports essential for effective administration. The scenario depicts a situation that is well suited for the implementation of distributed data processing (DDP). DDP is an arrangement of computing resources (hardware, data base, personnel) that is apportioned or shared to provide computing power where it is needed. The following sections discuss the historical development of distributed processing systems, the types and characteristics of these systems, and considerations for the development, implementation and management of distributed systems.

Development of Distributed Data Processing

Since 1954 when computers were first used in organizations for data processing, a wide variety of arrangements or configurations of computing power have evolved. Initially, computers were an expense only large organizations could afford. In most organizations computing power was concentrated in a single data processing department or center, commonly located in the central office. Users in the central office, as well as those in other offices, had to hand carry data to the data processing center, wait for the data to be keypunched and processed, and retrieve the results. To maximize the use of expensive computer time, data processing needs were accumulated or "batched" before processing, hence the term "batch processing." The economies of computing were clearly in favor of large centers. A standard rule of thumb was called "Grosch's Law" which said that, "a computer that costs twice as much as another one has four times the power" (Ralston & Meek, 1976:599). The logistical problems resulting from having to transport all data to the central site were eased by the late 1950's when communications of data over telephone lines allowed users to input and receive data at their organizational location (stage 2, remote job entry shown in Figure 1) but the central office still processed the data in a batch processing mode.

The mid-60's saw the development of time-sharing capabilities, which permitted multiple users to process their data simultaneously from a distance (Stage 3 in Figure 1). Computers with time-sharing capabilities could not only handle batch processing but also simultaneously operate in a different mode called interactive processing. In this mode, users are in direct contact with, or "on-line" with, the computer which could immediately respond to user commands. These new capabilities were significant, because they allowed users to operate far from the computing center while processing remained centralized. The next stage in computing came in the early 1970's with the capability to distribute the actual processing of data to multiple locations, hence the term "distributed data processing." The concept meant that computers could be located wherever necessary in an organization yet linked to form an integrated network of computers rather than isolated units. With the advent of low-cost minicomputer

Figure 1. Sequence of data processing forms



systems, Grosch's Law no longer applied. For example, today minicomputers which may cost 3 percent of a large main-frame computer can do much more than 3 percent of the work. In addition, hardware costs have continually declined as a percentage of total computing cost. Today hardware costs account for less than one third of the costs of running a typical data processing department. Other costs such as personnel and software have risen substantially and are becoming the dominant expense. The economies which favor distributed systems reach their limit, however, as the increase in complexity of a distributed environment leads to higher personnel and software development and maintenance costs.

It should be noted that all of the configurations in Figure 1 are in widespread use today. The best choice for a configuration depends on the particular situation. Furthermore, in many cases a completely centralized or decentralized environment (stage 1 in Figure 1) may not be the optimum computing solution, but tradition and inertia may have prevented any change. However, it is safe to say that the continual decrease in the cost of computing with minicomputers is making distributed computing an increasingly feasible option.

Types of Distributed Data Processing

A distributed computing environment can be defined as one in which some or all elements of the computing resource (data base, hardware, and personnel) are distributed in some fashion and to some degree. The initial approach was the distribution of hardware, i.e., dispersing the actual location of the computers in one of two major arrangements. In the first configuration, computers are organized in a hierarchy as shown in Figure 2, and the entire system is controlled by a large central computer. The other possibility is to arrange computers into a ring configuration as shown in Figure 3, in which the computers are autonomous but can communicate with any other as desired.

Most organizations are hierarchically structured and the first configuration is used because it fits operating procedures and the flow of information. Ring distributions are less common, being used when autonomous data processing centers need to communicate with each other. An example is the ARPANET system sponsored by the Advanced Research Projects Agency of the U.S. Department of Defense, which links together over 100 universities and research agencies (Kleinrock, 1973:305).

Hardware distribution provides just part of the total picture. Several other dimensions of data processing can be distributed in various degrees. Figures 2 and 3 also illustrate two other possible forms of distributed data processing, i.e., data base location distribution and personnel distribution.

The concept of DDP can be refined further by viewing data processing in an organization as occurring in numerous areas of activity which can be categorized under either control or execution (Buchanan & Linowes, 1980:143-153). Control activities are managerial in nature, such as personnel planning, budgeting, scheduling, staffing, and evaluation. Execution activities are divided into development activities (e.g., programming, systems analysis, user training) and operations (e.g., computer operations, telecommunications, and system maintenance). An organization should carefully select the appropriate degree of distribution of each activity.

Perhaps the best way to understand distributed data processing is to compare its characteristics with those of centralized data processing (see Figure 4).

Figure 2. Hierarchical Distribution

CENTRAL OFFICE

Typical Data Functions

- -payroll for total organization
- -collect, store, merge and report systemwide data
- -transmit systemwide data to all regional offices
- -provide reports which are standard across regions

REGIONAL OFFICES

Typical Data Functions

- -collect, store, merge and report data on regionwide operations
- -transmit regional data to central and local offices
- -provide reports which are standard across local offices of each region

LOCAL OFFICES

Typical Data Functions

- -collect, store, merge and report data on local operations, e.g., client data
- -transmit local data to region and suboffices
- -provide reports which are standard across suboffices of each local office

SUBOFFICES

Typical Data Function

- -collect, store and report suboffice data
- -transmit suboffice data to local offices

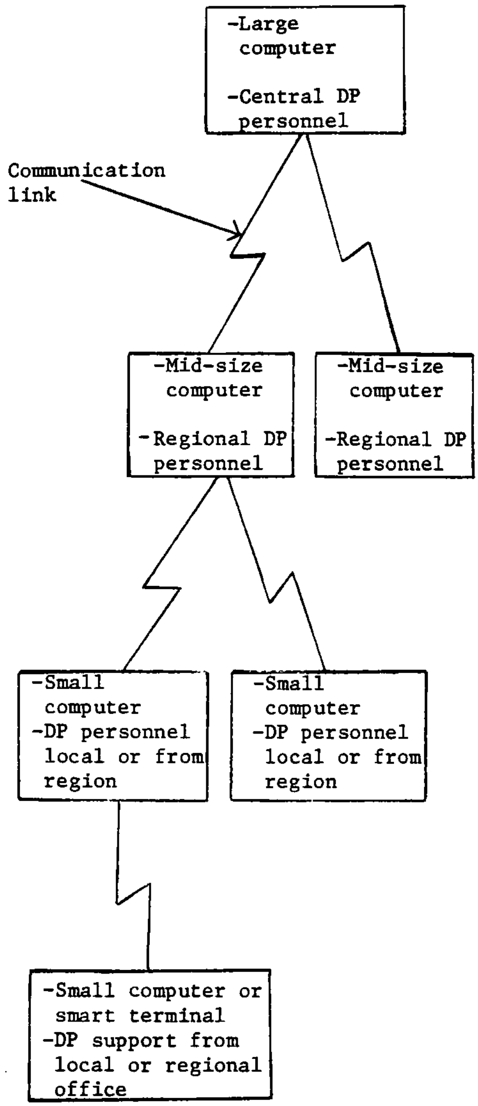
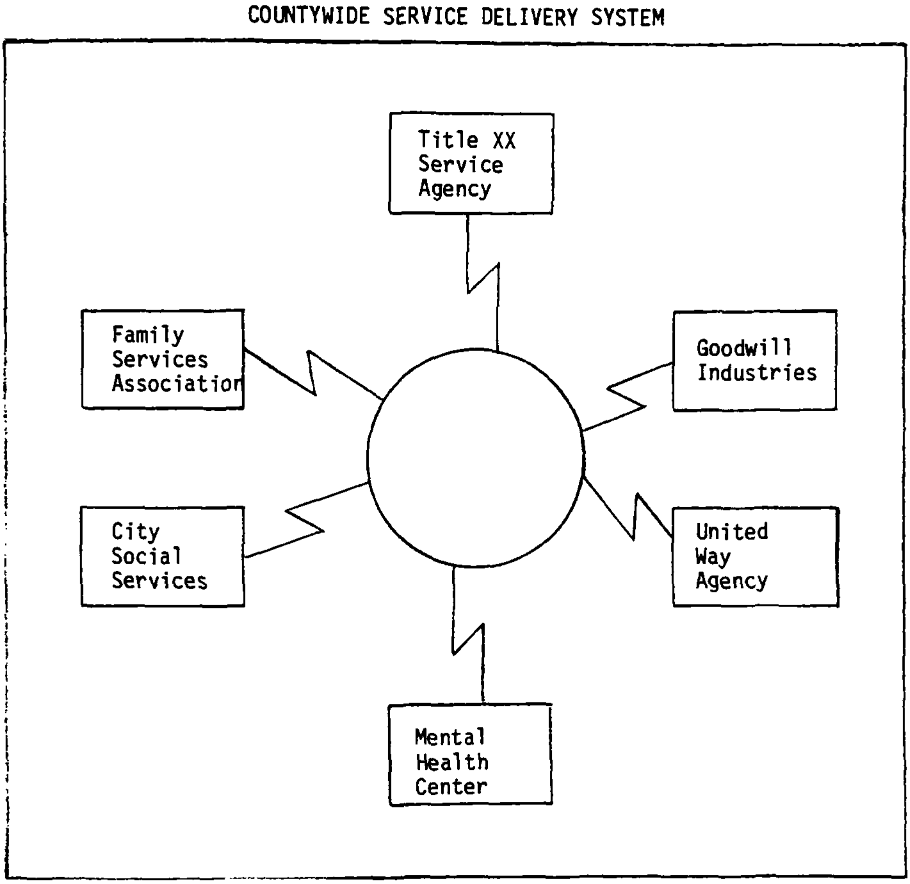


Figure 3. Ring Distribution



In a ring distribution, each agency owns its own computer. By connecting to the communications ring agencies can:

- -exchange service information
- -refer clients and client information
- -track clients through the service delivery system
- -share bibliographic information
- -share city or countrywide data for planning

Figure 4. Characteristics of centralized and distributed systems

CHARACTERISTICS OF A CENTRALIZED SYSTEM

Possible economies of scale in software, data, space, and expertise.

Capable of processing large complex applications.

Easier management and control of development and operations.

Greater capacity to expand hardware, storage, and input/output devices as CRTs and printers.

Higher telecommunications (long distance data transfer) costs throughout the system.

Higher complexity and therefore more chance for failure.

System failure disrupts entire organization.

Easier to protect data security and privacy.

Back up systems in case of failure are more costly.

Easier standardization and integration of data.

CHARACTERISTICS OF A DISTRIBUTED SYSTEM

Less overall hardware costs.

Redundancy exists in data acquisition and storage, personnel, software, programming and equipment.

Hardware is more readily available to user.

System failures can be handled by another component of the distributed system.

Possible to purchase improvements or enhancements in small low-cost increments.

Easier to tailor application to end user's needs.

More self-reliance in development and processing.

More flexible and adaptive to organizational change.

Encourages learning at distributed components.

Local control of data security and privacy.

Whether a characteristic of either processing mode is considered an advantage or disadvantage depends on such factors as organizational size, structure, managerial philosophy, current computing configuration, and so on.

Pitfalls of Distributed Data Processing

Several pitfalls can occur in a distributed data processing environment. These are discussed in the following sections.

Unplanned Growth

A distributed computer is typically purchased for a specific set of tasks. Nonetheless, a tendency exists to enlarge the scope of its applications, usually requiring extra hardware and software. If other possible problems are disregarded and only cost effectiveness considered, this creeping escalation may still be justified, because new ideas and new applications are part of any dynamic organization. However, many distributed computing centers do not have programmers or systems professionals attached to them. Running such centers is part of the job description of someone who is not a data processing specialist. If a point is reached when system size and complexity requires the hiring of full-time data processing personnel, then cost justification of new applications becomes much more difficult, because computing costs are currently declining at an annual rate of 15% whereas personnel costs are increasing at an annual rate of 6% (Emery, 1977). Recognizing this fact, many organizations have banned the hiring of data processing specialists at distributed processing centers. The result has been the hiring or transferring of personnel with other titles such as Administrative Assistant or Budget Specialist to spend their time operating the equipment, programming, or designing computer applications. The true cost of distributed computing often remains hidden because a variety of personnel may be devoting some of their time to computing.

Pioneering in Computer Technology

Most organizations should avoid implementing the latest developments in information technology. However, moving into a distributed environment places an organization on the "cutting edge" of distributed technology. Unless the distributed applications are relatively independent of each other and the central office, an organization will be involved in state-of-the-art techniques for determining what data to store at the various locations, how to locate needed data in the distributed system, how to synchronize the updating of data in the numerous distributed data files, and how to efficiently transfer data throughout the system while protecting clients' security and privacy.

Suboptimization

A third pitfall involves the situation in which the attempt to maximize the functioning of a subsystem detracts from the effectiveness or optimum functioning of the overall system. This situation is called suboptimization. A distributed system can produce economies of scale only if the subsystems do not suboptimize. For example, consider a state DDP system with one central office, ten regional offices with computing power, and five local offices which are too small to cost-effectively own a computer. The central office may process end-of-the-month reports for all fifteen sub-offices for $75.00, or five dollars a report. If regional offices can process this same report for four dollars, and reduce their own costs by doing so, the cost of the central office for processing the remaining five local reports will probably not decrease significantly because of high overhead costs. If it costs the central office $60.00 to process the remaining five reports, the fifteen reports now cost $100 for the state to produce. Thus, although ten regional offices are saving money from their own perspective, the cost to the state actually increases. A distributed computing environment can succeed only if subsystems agree to function suboptimally for the sake of the total system. This example also illustrates that the process used by the central data processing department for allocating computer costs to users must take into account such items as the cost of supporting a large central program library, sophisticated input/output hardware, and extensive user services.

These pitfalls can prove disastrous or their solutions can require an inordinate amount of time and effort for agencies moving into a distributed processing environment.

Implementing and Managing Distributed Systems

The previous section provides clues for organizations for implementing and managing distributed computing networks successfully. It may seem paradoxical at first, but the only route to a properly functioning distributed system is through proper control by a single computing unit. In order to avoid the protectionist attitude of many established central data processing groups, top management (via the computer steering committee) must establish its commitment to the level of distribution deemed appropriate and then provide the central staff with the necessary authority to implement this strategy. The concern here is not to create an adversarial relationship between distributed computing units and the central unit, but to administer the entire distributed network as one integrated computing facility. This will remove some of the advantages of distribution, e.g., complete independence of end-users, but the gains far outweigh the losses, because the distributed processing system can function as an integrated computer network under professional management and control rather than several centers which cannot interact with each other.

The functions of the central group must include the following:

1. Development of a master plan. The master plan should establish the basic structure of the distributed network for a two to four year time horizon. The structure of the organization is of major importance. Two major options exist, one to partition the agency based on geographic area (state office, regional offices, county offices, sub-offices), the other to partition the agency based upon functional areas (financial services, social services, children's services). In partitioning, high levels of communications and data processing activity should take place within the subsystems but much lower levels between subsystems. Links between distributed subsystems should be as simple and uncomplicated as possible. This can be achieved by isolating the most complex

activities at the distributed sites or by any other method which results in a structure with a high degree of independence of distributed centers from the rest of the system.

The thrust of the master plan will be reflected in data processing budgets of distributed components for each of the years under consideration. These budgets, will of course, be subject to approval by the steering committee.

If planning is done well, organizations can obtain far more for their computing outlays than with systems implemented in a haphazard fashion. Furthermore, a powerful and flexible management information system can only be made possible by building a common, integrated data base for the organization. Both these concepts--data base and integrated systems--would be impossible to achieve without strong central planning and control.

2. Control of the distributed system. Control is imperative if the entire system is to remain a coordinated entity, and if the entire distibuted computing network is to be administered as much as possible as one computer facility. However, the combination of geographic dispersion of facilities, possible ownership of these facilities by individual divisions or departments, and the political clout exercised by them makes this task more difficult than managing a central facility. Two primary tools for achieving control exist. First, a review can be made of computing budgets and acquisitions of hardware and software with approval required for major extensions to existing applications. Second, common standards can be developed that specify how documentation is to be written, how systems will communicate with each other, and how security and privacy will be safeguarded. The authority for approval is required because individual distributed units are invariably tempted to acquire hardware, software or personnel on their own. The standards are required to maintain compatibility among systems, ease systems maintenance, enable transfer of personnel and equipment, and accomplish data privacy and security objectives uniformly.

3. Provision of central computing services. Usually many groups require central computing services because they are too small to justify installing their own distributed systems, or they do not need such computing. Another category of applications requiring central services is that of large programs which cannot be run on distributed processors (usually minicomputers). An example of this is where support is given to planners and evaluators running large statistical programs. The central utility should also develop and operate applications which are common to all units. The best example of this is the payroll system. Other possibilities are a personnel skills file and bibliographic listing.

4. Provision of specialized technical support. Several specialized functions must be performed by the central group. Probably the most important of these functions is data base administration, if the organization has acquired a data base management system or it is planning to. In this case, the data base administrator will have to coordinate the usage of the central data base, which includes authorizing access and update, enforcing security provisions and other standards, maintaining the data dictionary, and optimally structuring the data base. Development and maintenance of the data communications network is inescapably a central responsibility. Like data base administration, this requires technical specialists with a high degree of sophistication. One of the serious problems in both areas is the scarcity of suitable personnel.

Some technical specialists can be shared with all distributed units, for example, systems programmers who can help to maintain the systems software throughout out the distributed network. In a similar fashion applications programmers can help in the design of applications throughout the distributed centers.

Another aspect of central support is the arrangement of training for all relevant personnel in systems-related matters. This training can take numerous forms ranging from in-house seminars and programs, to external courses, to university education. An extension of this service, which may help prevent the turnover of scarce data processing personnel, is the administration of a career development program for such personnel. Such a program would require tracking information such as employee skills, career paths, available training, and job openings (LaBelle, et.al., 1980:144-152).

Implications for Workers and Managers

In the ideal distributed system, the system's users are unconcerned about the location of the data they enter or request. All a user must know is what data are needed and in what format they are needed. A distributed system should act as one large system irrespective of where the data are entered, processed, or stored. For example, if a worker wanted to know if a client had previously been seen by the agency, the worker could type in the proper system entry codes followed by the client's name or identification number and receive immediate access to all information the security codes allowed. The information requested may actually be stored at several local, regional and state offices. When the worker's request was entered, the distributed system could search its directory of contents, determine where the relevant data resided, and retrieve, format and print the appropriate information for the user.

Although this ideal, well-connected distributed system will not exist for some time due to the complexities of the hardware and software required, distribution using present technology can do much to eliminate the problems of untimely, inaccurate, and irrelevant data which the introductory scenario typifies. An in-house computer could prepare all reports not needed by other offices of the system, thus giving local control over how rapidly the processing was completed and who had immediate access. Data from and to other levels could be transferred as necessary via telephone or mail. In addition, local workers could collect, store and report data unique to their office, thus making the system more relevant to the user. As local use of the system increased, the incentives for keeping accurate data would also increase (Schoech & Schkade, 1980:566-575).

One of the major problems in social welfare distributed systems is the standardization of service definitions if other than presently existing data are to be collected statewide. If every office is allowed to define its own data, then systemwide reports and management of the data collected systemwide will be impossible due to inconsistent definitions, reporting categories, timeframes, etc. Standardization of terminology, however, is an extremely difficult task requiring substantial time and effort by all data users. If incorrect data are collected or the right data collected improperly, the system will not and should not be used to support decision making. At present, the barriers to developing local computing power capable of processing more than existing data are not technological but concern the cost of software development and the categorization and systematization of worker activities.

Conclusion

Computing power is evolving just as other forms of power (e.g., mechanical power) have evolved. The first use of mechanical power involved a large centralized waterwheel or steam engine to run a factory. Today, factories use a mix of large and small motors to economically distribute mechanical power precisely where it is needed. The capacity to distribute computing power is just beginning. For large organizations, such as most social welfare systems, substantial work and change will be required before an efficient mix of computing power is obtained throughout the organization. This paper has discussed distribution, how to go about achieving this mix and some pitfalls to avoid. Just as centralized waterwheels and large motors were replaced, the centralized computer is becoming an obsolete technology for supporting the complex decision making that managers and workers in social welfare agencies must make.

References

Buchanan, J.R. and R.G. Linowes
1980 "Understanding distributed data processing." Harvard Business Review 58(4): 143-153.

Emery, J.
1977 Address to the International Federation of Information Processing Societies, Toronto. Proceedings of IFIPS, Geneva, Switzerland.

Kleinrock, L.
1976 Queuing Systems, Vol. II: Computer Applications. New York: Wiley.

Labelle, C.D., S. Kimball and L.J. Hellenack
1980 "Solving the turnover problem." Datamation 26(4): 144-152.

Ralston, A. and C.L. Meek (eds.)
1976 Encyclopedia of Computer Science. New York: Petrocelli.

Schoech, D.J. and L. L. Schkade
1980 "Computers helping caseworkers: decision support systems." Child Welfare 59(9): 566-575.

SELECTED READING ON DISTRIBUTED PROCESSING

Auerbach Editorial Staff. Distributed processing management. (Auerbach Information Management Series), N.J., Pennsauken: Auerbach Publishers, 1980.

Donaldson, H. Design of distributed data processing systems. New York: Wiley and Sons, 1979.

French, J.C. Selecting a distributed system, Datamation, 1981, 27(2), 67-69.

Lecht, C. The waves of change. Computerworld, 1977 11(25), 23+.

Lorin, H. DDP: how to fail. Datamation, 1981 27(2),60-65.

Snyders, J. Those belated distributed DBMSs. Computer Decisions, 1981 13(2), 76-96.

Shirey, R.W. (Series of four articles on distributed computing), Computerworld, 1980, 14(41) (42) (43) (44).

Withington, F. Coping with computer proliferation. Harvard Business Review, 1980, 58(4), 152-164.