

THREE REASONS FOR THE UNDERUTILIZATION OF SOCIAL SCIENCE DATA SERVICES IN THE INFORMATION AGE

ALICE ROBBIN

DATA AND PROGRAM LIBRARY SERVICE
UNIVERSITY OF WISCONSIN-MADISON

Societies everywhere are being affected by the new information technologies. In addition, they have become increasingly dependent on statistical information for making important public policy decisions. Social science data services, a direct result of the new technologies, have been established to provide easier access to computerized statistical information. One would expect therefore to have seen over the last 15 years a great many data services established throughout institutions of higher learning and government. Yet these data services are few and the ones that exist, underutilized. There are obviously many reasons for their underutilization.

Today, I will address three reasons which contribute to the current situation. Poor quality data impede good decision making and research. Lack of coordination and planning of the statistical information system make it very difficult to produce, locate and retrieve data. New information technologies are modifying our societies. But social scientists are not directing enough attention to how society is being altered and we lack appropriate models and data. My concluding remarks suggest a number of ways that social scientists can contribute to improving the current situation.

Social science data archives and services, like their predecessor libraries and archives of print documents and film, represent one component of a society's institutional memory. The underlying philosophy of preservation and access holds that transfer of the data collections from their producers to these data centers greatly increases the return on the original public and private investment.

This paper was delivered at the 1981 IFDO/IASSIST conference in Grenoble. The author gratefully acknowledges helpful comments on earlier versions from Thomas Flory, Nancy McManus, and Richard C. Roistacher.

Most of these centers were established before their national archives created machine-readable divisions. Although these centers have not been designated official repositories for government records, governments have turned to them for assistance in retrieving government data files. As recent experiences in several countries demonstrate, more government data producers are delegating archival responsibilities to university data repositories in recognition that government cannot preserve and maintain its own records.

As society's problems have grown more complex, statistics have become more important to effective decisionmaking. Not only do policymakers face increasingly complex issues, but many problems now interact with one another (12,136).

The resources of data centers, for holding historical collections of data and for generating new ones, are essential if national policy decisions are to be made in a more rational manner. Existing administrative records systems, used for secondary analysis or linked to new data collection activities, provide a means for responding efficiently to new policy questions.

Data services are also important for cumulative social science research activities. Common access creates a "commonality of research among widely separated scholars" (9,411). The data archive acts as a scientific laboratory which encourages the sharing of data, multi-disciplinary exploitation of evidence, and "multiple and complex analytic applications" (5,393). The data center makes a pedagogical contribution by allowing the student to participate in scientific inquiry, developing problem-solving techniques and behavior like those of students in the natural sciences. A recently completed study of factors influencing the sharing of computer-based resources for higher education and research shows a direct connection between utilization and sharing. It suggests that the "seemingly indirect attempts to broaden 'computer literacy' and computer use might have systemic effects on the level and nature of computer-based sharing" (8,4.44).

Less obviously, the data archive plays a role as an agent for assessment of information transfer activities. It offers administrators and researchers the opportunity to assess the technical, administrative, economic, and policy issues related to standards of data quality, documentation, access, and distribution.

Nevertheless, 33 years after the Roper Center at Williams College in Massachusetts and 20 years after the establishment of the Steinmetz Archives in Amsterdam, the Zentralarchiv für Empirische Sozialforschung at the University of Cologne, and the Inter-university Consortium for Political and Social Research at the University of Michigan, no more than 50-odd data services exist throughout the world, almost all university based. National governments have been slow to accept the idea that data services play an important role in information policy development.

Information technologies and services produced and offered by the private-for-profit sector are beginning to dominate access channels.

Why are there now so few social science data archives and why do they appear to be underutilized? That they have been is due to a wide array of reasons. Rather than providing an inventory of these reasons, I will address the complex and interdependent issues of the quality of statistical data, factors responsible for the lack of coordination of data resources, and the need to make social science more relevant to policy choices.

Most of my remarks have been stated in one form or another during the last five years in many countries. I address the creation of statistical data and administrative records produced by government because it is a major provider of the data resources which social scientists use. And I expect that in the future, government's influence on statistical data production will determine even more how the social scientific community conducts itself.

My remarks about the role of social science in an Information Age have been influenced by recent political events, in which many questions have been raised about the relevance of social science. I believe that relevance implies and requires philosophical reflection. Relevance requires use of theoretical perspectives about human and social interests. Relevance requires new models which integrate our natural and social worlds with scientific and technological discoveries. My recommendations for improving data quality, planning, and coordination should be understood as two aspects of the larger philosophical and moral dilemmas which we confront. Thus, the last part of my address reflects on some of the questions social scientists must seek to answer as they confront social changes which are the result of new information technologies.

II. Problems of Data Quality and of Coordination and Planning

A. Data Quality

Dissatisfaction with the quality of data is widespread throughout the scientific community and government, although enormous strides have been made to improve measurement. David R. Lidd, Jr., director of the Office of Standard Reference Data at the U.S. National Bureau of Standards, recently wrote "that a considerable amount of information in such archives is erroneous". He cited almost 200 reported measurements of the heat conductivity of copper--"a range of values so great that most of the data are clearly off the mark" (12). Publications of social and science indicators, on which many projections in the United States are based, contain obvious statistical errors--obvious, that is, once the data are examined--and inadequate information on sectors of the society which we know are undergoing rapid changes. These errors are due in part to inadequate sampling frames and improper methodological tools applied to data gathering and analysis.

Three factors that influence the quality of statistical and other data and their analytic potential are demand (or user requirements), supply (or the resources of the system), and structural or environmental conditions.

User Requirements. A recently published White Paper on the U.S. statistical system notes that "the complexity and urgency of issues facing policymakers often leads them to demand more data and more timely data, with little regard for quality" (11,164). Policymakers tend to be uncritical about the quality of the data they use; social scientists only somewhat less so. The immediate demands for completing the administrative function, a budgetary horizon of one to two years, and legislative demands for information for modifying policy impede the necessary gestation period for designing and gathering data. Political ends influence the quality of data. ". . . Some of the most important statistics are held hostage to political ends by their visible and direct use in politically important decisions which allocate [national] resources" (4,204).

Resources for Maintenance and Improvement of Quality. At least in the U.S., there has been no thorough government-wide review of classification standards for statisticians for about three decades. Professional training in data handling is received (or not received, as the case may be) on the job, with little influence by non-governmental sources of expertise. The social science community, which has discovered many useful tools for improving data quality, has little opportunity for interaction with the governmental data producer and statistician. This interaction is not encouraged by government and the university organization nor by attitudes of the government administrator or academician. Civil servants' opportunities for career development and participation in conferences such as this one are limited.

The White Paper offers other explanations. Budgets for statistical programs and projects do not include resources for internal and/or external measurement of quality. Funds are seldom provided for methodological research to improve quality, except where there are clear indications of serious deficiencies. Such deficiencies may not become obvious until the effects of poor policy decisions are felt. Political bodies are then moved to apply remedies (which rarely reflect the underlying systemic problem). Little attention is given to the basic design of surveys, evaluation studies, program experiments, and data bases developed for policy analysis. Competitive procurement activities (contracts, for example) seldom receive adequate technical review, and selection panels often lack the technical skills to make an informed judgment (11).

A 1978 study by the U.S. General Accounting Office of federally-sponsored attitude and opinion surveys found serious technical flaws which limited the usefulness of the results in all five surveys which were reviewed in detail. The GAO concluded that "better guidance and controls were needed to improve Federal surveys of attitudes and opinions" (11,162). Another study, sponsored by the American Statistical Association

and funded by the U.S. National Science Foundation, evaluated 26 sample surveys conducted in 1975 and found that 15 of the 26 surveys had serious technical flaws. All but two of the 26 federally sponsored surveys were conducted under contract by universities or other private survey research organizations (1).

Structural factors affecting quality. Increasingly, statistical services are being procured from outside the government under contract. Agencies often have funds to acquire these statistical services, but no budget to develop staff and inhouse organs to build services and decide on technical specifications and selection. Operations which include data collection by other units of government are notoriously difficult to monitor and to standardize. For example, a large portion of the data collection activities conducted under the auspices of the intergovernmental Cooperative Health Statistics System program in the U.S. is being eliminated; quality control was cited as a major factor in this decision (13). Producers outside government are typically unaware of the uses to which their data will be put, or of the utility of the data they provide or of the administrative needs of an agency. Analysts are often unaware of important limitations of data because technical standards of data description have not been instituted by government agencies.

Restrictions on interagency sharing often result in the lack of comparability in data produced by different agencies. Such restrictions sometimes result in failure to fully exploit expensive data bases. Although policy may require linkages of materials gathered in several agencies and from several records series, legal procedural, and operational mechanisms to provide linkage are few and far between(2).

B. Lack of Coordination and Planning

Poor information management practices applied to statistical and administrative records and the internal organization of bureaucracy are in part responsible for difficulties in accessing records. These problems have led "to a growing incidence of overlap, duplication, mismatch and gaps in data and analysis, and increasingly complex problems of access by users and statistical agencies to various Federal data" (11,143).

Nora and Minc give three examples of this kind of compartmentalized development in France.

Hospitals have developed systems for billing medical expenditures and hospital-stay expenditures without collaborating with Social Security. Within Social Security itself, compartmentalization into three branches, each with its own data processing centers, has led to manual retrieval of data produced by the computers of the other branches. As a result of the present departmental separation [they write before various reorganizations within the Mitterand government], the Direction Générale

des Impôts and the Direction de l'Aménagement Foncier et de l'Urbanisme (land development and urban affairs) has each established a land use data bank, the former for tax purposes, the latter for development purposes. The legal definitions and the types of information differ. Nevertheless, there are broad common areas, but nobody worries about them. In addition to the waste, the establishment of these two data banks prolongs administrative isolation. Strengthened by this investment, both administrations are prepared to resist attempts at rapprochement (10,115).

Within the U.S. government, the Federal Trade Commission in its quarterly financial reports asks for data which are available in quarterly filings with the Securities and Exchange Commission. And there are currently three duplicate mortgagee interest surveys (11,149). In Wisconsin, the Department of Public Instruction refuses to turn over computerized records that the Department of Revenue needs for statistical analyses and modeling. The Department of Revenue is forced to collect this information manually if it is to perform its work in a timely way.

The application of data processing technologies has been uneven throughout government, and as Nora and Minc note, although "penetration has been extremely rapid," it has "taken place in uneven ways, strengthening barriers, immobilizing the structures that it penetrates for a long time" (10,112). They note that

in the majority of cases, each department acquires data processing capabilities without worrying about the possible difficulties that its plan may cause elsewhere, and especially without measuring the "synergistic" effects that better coordination with other departments might have produced (10,112).

The high rate of change in administrative data processing has resulted in a phenomenon that could be called input without throughput. Delays in the implementation of data base management systems, complications in electronic data entry systems, pressures to maintain routine administration in the face of high staff turnover in data processing, and the imposition of computer technology on organizations designed for manual systems have created serious bottlenecks in routine administration. Procurement policies emphasize centralization and are costly and a serious impediment to acquiring the most economical and efficient technology available. Little attention is given to identifying areas where decentralization of the information system would improve an agency's capabilities. On the other hand, administrators have few possibilities and little incentive to improve coordination because statutes delimit an agency's mission.

Even when research access to identifiable information is not in question, attention has not been given to maintenance and preservation of machine-readable records. Constraints on administrative activity

tend to reduce incentives for "backward" looks, those that would require that records be maintained and preserved. The resulting costs can be very high. For example, efforts now underway to create public use samples from microfilmed versions of the 1940 and 1950 U.S. Censuses of Population are to cost \$8 million. Much of that information was on punch cards at one time. Records managers and archivists do not usually participate in decisions about retaining and destroying computerized records. As a result, computerized records are not integrated into records management practices.

Records managers leave decisions about retention to those with programmatic responsibility and concern themselves with managing paper and microfilm records. Records and computer centers see themselves as repositories for magnetic tape, with responsibility for decisions about tape maintenance left in the hands of an agency. Individual analysts retain information on the contents of files for which they have programmatic responsibility. Data processors are often the only persons knowledgeable as to format and physical attributes of computerized records. Documentation for MRR may not exist or may be scattered among the various agency personnel responsible for the different aspects of MRR. Valuable data are routinely erased and the tapes are reused when tape shortages occur, often without prior systematic review.

III. Society and the New Information Technologies

The emerging information technologies are already altering the nature of our society and affecting existing political, economic, and social institutions and values. Data processing is accelerating production,

with less but more effective work and jobs very different from those imposed by industrial life. This change has already begun: a great decrease in the labor force in the primary and secondary sectors, an increase in the services, and above all, a multiplication of activities in which information is the raw material (10,126).

Already, computerization of formerly manually performed tasks is rendering the semi-skilled and unskilled worker unemployable. Robots are beginning to replace humans, performing certain tasks more efficiently and increasing industrial productivity. However, not only the unskilled or low-skilled are being replaced. The introduction of automation is affecting highly skilled technical workers. For example, although more than 12,000 air traffic controllers walked off their jobs in the United States, air traffic was only partially reduced because computers assisted in air traffic decision making. In the opinion of some, computers were used as a strike-breaking tool(3). The Federal Aviation Administration hopes within 10 years to have computerized en route air control to such an extent that at least 50% fewer controllers will be needed and those that will be needed will be computer managers(6).

Economic changes will be accompanied by a change in the structure of organizations and by fluctuations in attitudes toward work. As numerous examples have demonstrated, the new technologies related to automation and data processing can flourish in small as well as large organizations. The psychological and social bonds that were created by the work place and that fostered worker solidarity will weaken as automation enforces isolation.

Monetary and other rewards will go increasingly to those who have the means to produce and manipulate the technology, creating new elite structures and placing political decision-making in the hands of technicians. As Duncan McRae has noted, the "risk of technocracy lies in the possibility of uncontrolled power held by an elite and devoted to special values and interests rather than to the general welfare" (7,45-46).

IV. Recommendations

In what ways can social scientists contribute to improving the present environment of the information system? The information system in which statistical data production and analysis take place is highly complex and dependent on new technologies. It requires expertise from many disciplines and specializations. It requires modifications in the institutional framework in order to cope effectively with societal change and to anticipate unexpected policy and political demands.

The social scientist and policymaker have many common interests. They have a great deal to gain by cooperating, to improve the quality of data, coordination and planning, and access to computerized records. Governments must use available expertise "in data collection and analysis activities, starting at the design stage, and continuing through to evaluation of how results are used" (11,166). Social scientists can contribute through methodological research in measurement of errors to improving collection methods and to improving the presentation of information about methodology structure and other limitations of the data products and analyses. The results of methodological research must then be widely disseminated so that they can be evaluated, criticized, and competing methods proposed if necessary.

We must be concerned with creating an integrated output and with producing cross-cutting analyses over a wide range of issues. Social scientists can assist in substantive integration activities, by developing standard concepts, definitions, classifications, survey frames, and procedures, and by monitoring and promoting their utilization by government and by the private sector. Social scientists can assist in developing a "consistent conceptual framework or model based on behavioral relationships in various disciplines" (11,172).

There needs to be increased use of administrative records to produce statistics and to respond to public policy questions. Public use samples should be drawn from administrative records. Administrators should be

made to produce public use files and to coordinate record linkage and analyses. Through their activities, social scientists can promote record linking at the microlevel and demonstrate ways in which the data's analytic potential can be enhanced. (It is important to note, by way of illustration, that social scientists and government officials in Germany have been meeting to discuss the creation of public use samples. This meeting should be emulated by other countries.)

Some of the problems of use of social science methodological and policy research can be traced to the fact that researchers are not part of the policy formation activities of government. If social researchers are to play a greater role in social policy formation and are to increase utilization of their research, there must be a higher rate of communication between researchers and policymakers. This communication is more successful if social scientists participate in internal organizational decisions (14). Social scientists must make a concerted effort to involve themselves in these decisions. Involvement in the internal decision-making process will indirectly improve the quality of civil servants' activities and directly improve utilization of their research and policy recommendations.

With administrators and policymakers, social researchers can assess research needs and examine the relation of the statistical system to research activities outside the government. They can apply their training in organizational theory and public administration to improving information management activities in government. Indeed, some of these very activities are already underway in Italy, Norway, Germany, the United States, and Great Britain.

Closer ties between data producers and analysts will result

in data that are more relevant to policy issues and will also improve the quality of both data and analyses. Producers of data will have more direct feedback on quality from major users of data . . . Users will come to have a better understanding of the operational problems of collecting and processing data, and will design and perform their analyses with a better understanding of the limitations of the data (11,168).

What should be the role of social science in an Information Age? This is a much more difficult question than the one which asks what knowledge should be applied and how? Let me identify only a few salient public policy issues that form part of an agenda for information technologies-related social research and training.

(1) Society will require a decreasing amount of work. Will work as a value lose its importance? How will the remaining work be distributed? What educational and job training programs will be needed, ones that are more compatible with the requirements of the post-industrial and information age? If the number of hours of leisure time is

increased, what social and psychological changes will occur; what changes will be necessary?

(2) New organizational structures are evolving and, increasingly, innovation takes place and new products develop in small units. What should be the role of the state in reorganizing the production structures? How do we design tax policies and write administrative regulations to provide incentives for industrial and university research and development, to foster innovation and risk-taking in the highly productive information technologies? If basic research outside industry is a prerequisite for innovation and continuing productivity, are the existing models of research in a more decentralized fashion, along the lines of the U.S. model, or research in the Colbertist tradition any longer relevant; or is some mix more appropriate to optimize available resources and to encourage innovation?

(3) Critical shortages of trained scientific and technical personnel are beginning to be felt. In what ways can we improve the quality of our science and social indicators to reflect the current situation? How can we estimate the impact of these shortages on the economy and on a nation's productivity? What roles should the state and the private sector play in ameliorating these conditions? If university budgets continue to experience serious erosion, how will a nation's productivity and general welfare be affected? Yet, if attention is turned only to reducing these shortages, do we risk neglecting the education of the "well-informed citizen" who is necessary for democratic control of technical decisions? Do we thus accelerate the creation of a society which is, to quote Shils, "victim of the parochial preoccupations of specialized technical experts"? [in McRae]. If we emphasize scientific knowledge to the detriment of valuative discourse will we neglect the education of both the scientist and the consumer of technology?

(4) The design industry and regulatory arms of the state have been preoccupied with hardware systems, with minimal consideration of human factors and a disregard for worker participation. The accident at Three Mile Island nuclear power facility on March 28, 1979, dramatically illustrates the failure to integrate the reactor operator into the system. The Kemeny Commission pointed to the mutual isolation of the operator and equipment in the highly complex sociotechnical system as a root cause of the accident (15, 57). The social scientist Malcolm Brooks observed that the events were a direct function of the electro-mechanical system design and detail (15, 58).

In what ways can we improve the man-machine interface in order to reduce isolation and alienation? If it is necessary to modify the work environment, in what ways? Are our theories of participatory democracy relevant to the emergence of new environments based on information technologies? (Is the model of industrial democracy relevant in a post-Industrial Information Society?) Can the new information technologies and new sources of knowledge enhance autonomy and responsibility, make possible mastery of the natural and social world, and emancipate rather than imprison us?

(5) Instrumental reason has spread to many areas of social life and there is an increasing tendency to define practical problems as technical issues. Will technocratic domination erode the institutional framework of society? What value system will it dictate? Will the technical values of efficiency and economy dominate the selection of means for realizing social goals?

(6) The ability to communicate has always been the purview of the educated and dominant classes. Will standardization of access vocabularies affect language and syntax and authority structures? If language will be of a different nature, simplified, to reduce communication costs, will we then sacrifice part of the content? What will occur when the essential meaning of messages related to daily life becomes available to anybody? Will new communication structures create more open and accountable authority structures? Do they offer the potential of transforming the state into one more easily supervised by the "public"?

(7) The cultural model of a society also depends on its memory, control of which largely conditions the hierarchy of power. Will access to infinitely greater sources of information entail basic social changes and affect the social structures by modifying the procedures for acquiring knowledge? (10, 313). How will data banks restructure knowledge? How much social control will be exercised by the producers of data banks?

To understand the nature and direction of technological change demands a vigorous and sustained program of social research related to information technologies. The frameworks of the social science disciplines and social thought can help us in orienting our discourse and directing it to problems of action and choice. New information bases and new knowledge can improve political choices in an increasingly technological society. They can assist social groups to transform society, to use new resources effectively and to their benefit, and to create control mechanisms for the New Information Order. This effort requires engaging and appropriating competing traditions of philosophy and social thought, new philosophical approaches and different methodologies, and creativity and innovation unfettered by the narrow confines of the empirical sciences.

References

- (1) Bailer, B.A. and Lanphier, C.M. (1978) Development of Survey Methods to Assess Survey Practices. Washington, D.C.: American Statistical Association.
- (2) David, M. and Robbin, A. (1981) The Great Rift Between Administrative Records and Knowledge Created through Secondary Analysis. Review of Public Data Use. (forthcoming)
- (3) FAA withheld air control system until strike: Aspin. Capital Times, August 24, 1981, 23.

- (4) Federal Statistical System Project Staff. (1980) Improving the Federal Statistical System: Report of the President's Reorganization Project for the Federal Statistical System. Statistical Reporter, May, 1980, 197-212.
- (5) Hofferbert, R.I. and Clubb, J.M. (1976) Introduction. American Behavioral Scientist, 19 (4), 381-386.
- (6) Kolata, G.B. (1981) FAA Plans to Automate Air Traffic Control. Science, 213, 21 August, 845-46.
- (7) McRae, D. (1976) The Social Function of Social Science. New Haven, CT: Yale University Press.
- (8) Mebane, D.D. and Mebane, R.M. (1981) Factors Influencing the Sharing of Computer-Based Resources for Higher Education and Research. Princeton, NJ: EDUCOM.
- (9) Miller, W.E. (1976) The Less Obvious Functions of Archiving Survey Research Data. American Behavioral Scientist, 19(4), 409-418.
- (10) Nora, S. and Minc, A. (1976) The Computerization of Society (L'informatisation de la société). Cambridge, MA: MIT Press (translation).
- (11) The President's Reorganization Project for the Federal Statistical System. (1981) Improving the Federal Statistical System: Issues and Options. Statistical Reporter, February 1981, 133-221. ("White Paper")
- (12) Sullivan, W. (1981) Data Services Map the Way in Labyrinths of Information. The New York Times, August 23, 1981.
- (13) U.S. Department of Health and Human Services. (1980) Directions for the '80s. Final Report of the Panel to Evaluate the Cooperative Health Statistics System. Washington, D.C. National Center for Health Statistics. (DHHS Publication No. (PHS) 80-1204).
- (14) Van de Vall, M. and Bolas, C.A. (1981) External vs. Internal Social Policy Researchers. Knowledge: Creation, Diffusion, Utilization, 2(4), 461-482.
- (15) Wolf, C.P. (1979) The Accident at Three Mile Island: Social Science Perspectives. Items, 33(3/4), December 1979, 56-61.